OPEN ACCESS

ORIGINAL RESEARCH

Reasons for and time to retraction of genetics articles published between 1970 and 2018

Rafael Dal-Ré ,[1] Carmen Ayuso[2,3]

[1]Epidemiology Unit, Health Research Institute-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid, (IIS-FJD, UAM), Madrid, Spain
[2]Genetics and Genomics Department, Health Research Institute-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid, (IIS-FJD, UAM), Madrid, Spain
[3]Centro de Investigacion Biomedica en Red de Enfermedades Raras(CIBERER), ISCIII, Madrid, Spain

Correspondence to
Dr Rafael Dal-Ré, Unidad de Epidemiología, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz, Hospital Universitario, Avda Reyes Católicos 2, Madrid 28040, Spain;
rafael.dalre@quironsalud.es

Check for updates

ABSTRACT

Introduction Between 0.02% and 0.04% of articles are retracted. We aim to: (a) describe the reasons for retraction of genetics articles and the time elapsed between the publication of an article and that of the retraction notice because of research misconduct (ie, fabrication, falsification, plagiarism); and (b) compare all these variables between retracted medical genetics (MG) and non-medical genetics (NMG) articles.

Methods All retracted genetics articles published between 1970 and 2018 were retrieved from the Retraction Watch database. The reasons for retraction were fabrication/falsification, plagiarism, duplication, unreliability, and authorship issues. Articles subject to investigation by company/institution, journal, US Office for Research Integrity or third party were also retrieved.

Results 1582 retracted genetics articles (MG, n=690; NMG, n=892) were identified . Research misconduct and duplication were involved in 33% and 24% of retracted papers, respectively; 37% were subject to investigation. Only 0.8% of articles involved both fabrication/falsification and plagiarism. In this century the incidence of both plagiarism and duplication increased statistically significantly in genetics retracted articles; conversely, fabrication/falsification was significantly reduced. Time to retraction due to scientific misconduct was statistically significantly shorter in the period 2006–2018 compared with 1970–2000. Fabrication/falsification was statistically significantly more common in NMG (28%) than in MG (19%) articles. MG articles were significantly more frequently investigated (45%) than NMG articles (31%). Time to retraction of articles due to fabrication/falsification was significantly shorter for MG (mean 4.7 years) than for NMG (mean 6.4 years) articles; no differences for plagiarism (mean 2.3 years) were found. The USA (mainly NMG articles) and China (mainly MG articles) accounted for the largest number of retracted articles.

Conclusion Genetics is a discipline with a high article retraction rate (estimated retraction rate 0.15%). Fabrication/falsification and plagiarism were almost mutually exclusive reasons for article retraction. Retracted MG articles were more frequently subject to investigation than NMG articles. Retracted articles due to fabrication/falsification required 2.0–2.8 times longer to retract than when plagiarism was involved.

corrections, expressions of concern and retractions of published articles. The latter concerns findings that result in loss of confidence in the study findings, in the conduct of the study or in the publication process of the manuscript. In biomedical and life sciences publications, retracted articles accounted for 0.02% of all published research articles indexed by PubMed as of May 2012.[1] Research misconduct (ie, fabrication, falsification, plagiarism)[2] accounted for 53%, whereas duplication and error were responsible for 14% and 21%, respectively; the USA was the country with the highest number of retracted papers due to research misconduct.[1]

In 2010 the Retraction Watch blog started to comment on retraction notices.[3] A database was initiated to host all the information regarding retractions, corrections and expressions of concern. The Retraction Watch database (RWdb) has been used for research for some time, but it was only opened to the public in October 2018. As of 31 December 2018, the RWdb had hosted 18 649 retracted articles from all disciplines.[4] It is considered to be the largest and most comprehensive database of retracted articles.[5]

Although genetics has experienced an unprecedented development in this century, it has a long history of scientific research. The RWdb has hosted retraction articles on genetics since 1977, starting with four retracted articles by RJ Gullis and CE Rowe—for yielding no reproducible results[4]—published in 1975. Genetics is the subject of research of any life science discipline. Along with biology, genetics could be the discipline that involves the greatest variety of researchers—with the implication of widely differing ethical approaches to research integrity. To the best of our knowledge, no research has been conducted that specifically aims to describe the main features of retracted genetics articles. Since genetics encompasses both medical-related disciplines and non-medical-related disciplines such as anthropology, biology or botany, a distinction between features of retracted medical genetics (MG) and retracted non-medical genetics (NMG) articles was warranted. The aims of this study were: (a) to describe the reasons for retraction of genetics articles and the time elapsed between the publication of the original article and that of the retraction notice (or time to retraction) because of research misconduct (ie, fabrication, falsification, plagiarism); and (b) to compare all these variables and the time needed for retraction of those cases of research misconduct investigated by the US Office of Research Integrity (ORI), between retracted MG and NMG articles.

INTRODUCTION

Journal editors are responsible for maintaining the integrity of published articles. To this end, in addition to ensuring the standards of the editorial process before publication, they can publish

BMJ

MATERIAL AND METHODS

On 14–16 January 2019 a search was conducted on the RWdb for articles published between 1 January 1970 and 31 December 2018. Boolean operators used were 'OR', 'AND' and 'AND NOT'. The following descriptors were applied. Nature of notice: 'retraction'. Article types: 'case report', 'clinical study', 'commentary/editorial', 'letter', 'meta−analysis', 'research article', and 'review article', using the Boolean operator 'OR'. This strategy allowed the exclusion of other types of article such as, for example, 'conference/abstract/paper', 'dissertation/thesis', and 'government publication'. Subjects: 'genetics'; MG was searched as 'genetics AND medicine', whereas NMG was searched as 'genetics AND NOT medicine'. Within the 31 subjects/specialties covered by 'medicine', the following were excluded: complementary medicine, dentistry and nursing. Then, as predefined by the RWdb, the following 'reasons for retraction' were searched: fabrication/falsification, plagiarism, duplication, unreliability, grave issues with authorship, results not reproducible, contamination of cell lines/materials/reagents, conflicts of interest, fake peer review, and error by journal/publisher. Of these 'reasons for retraction', only those containing a minimum of 50 cases in both datasets—MG and NMG—were considered, to allow for establishing reasonable comparisons. We also searched for those cases that were subject to civil or criminal proceedings, that provided limited or no information or that were subject to investigation (by 'company/institution', 'journal/publisher', 'ORI' or 'third party').

The following information was retrieved from each retracted article: date of publication of the original article and that of the retraction notice, PubMed ID (or DOI) of the original paper, and country (or countries) of origin of the investigating teams, with special interest in the seven countries (China, Germany, India, Japan, South Korea, UK, USA) with the largest number of biomedical and life sciences retracted papers[1]. All data were retrieved and checked by RDR; a random sample (by means of a random integer generator; https://www.random.org/) of 25% of all the data was checked for consistency by CA, and no discrepancies were found.

Statistical analysis

Retracted genetics articles

Descriptive statistics of the reasons for retraction and investigation, and time to retraction of research misconduct retracted articles, were calculated. To ascertain if there was any temporal trend in the reasons for retraction and in the time to retraction, logistic and linear regression analyses were performed by dividing the total study period into five time periods: 1970–2000, 2001–2005, 2006–2010, 2011–2015, and 2016–2018. This allowed comparisons to be made between the last four periods of time considered and the first one (1970–2000), which was used as a reference. In the logistic regression analysis, results obtained for the last period (2016–2018) should be taken with caution because the time for retraction was too short in comparison with those of the three previous time periods considered; similarly, in the linear regression analysis, results obtained for some time periods should be viewed with caution since the number of cases was too small in comparison with the other time periods considered. All tests were two-sided; $p < 0.05$ was considered to be statistically significant.

Retracted MG and NMG articles

The descriptive statistics of the reasons for retraction and investigation, and time to retraction of research misconduct articles

of both MG and NMG, were calculated. Odds ratios were calculated in each of the five study time periods (1970–2000, 2001–2005, 2006–2010, 2011–2015, 2016–2018). χ^2 test was used to ascertain if differences in the percentage of the reasons for retraction and investigation existed between both datasets. Bonferroni correction was made for multiple testing. All tests were two-sided; $p < 0.004$ was considered statistically significant.

Mann-Whitney U test was used to compare time to retraction of MG and NMG articles where fabrication/falsification or plagiarism were involved, over the five time periods. Bonferroni correction was made for multiple testing. All tests were two-sided; $p < 0.0125$ was considered statistically significant.

Finally, the descriptive statistics of time to retraction of investigations conducted by the ORI for both retracted MG and NMG articles were calculated and a comparison between them was performed using the Mann-Whitney U test. The test was two-sided; $p < 0.05$ was considered statistically significant.

Statistical analyses were performed using R.3.5.1 (R foundation for Statistical Computing, Vienna, Austria).

RESULTS

Retracted genetics articles

Following our search strategy (online supplementary information−1), there were 1584 retracted articles on genetics; 692 (44%) on MG—although two were excluded for being on complementary medicine and dentistry—and 892 (56%) on NMG. Table 1 shows all the reasons for retraction and investigation for all retracted genetics articles (online supplementary information−2). Research misconduct and duplication were involved in 33.2% (525/1582) and 24.3% of retracted articles, respectively. Thirty-seven per cent (587/1582) of genetic retracted articles were subject to investigation. Of the 10 reasons for retraction, only five met the minimum number of 50 article retractions in both datasets that allowed for inclusion in the analysis. Among the seven countries assessed, and related to the five reasons for retraction and investigation, the USA (n=526) and China (n=509) were the origin of the authors of the majority of retracted articles; South Korea (n=64) and the UK (n=69) were the countries with the lowest number of retracted papers

Temporal trends by means of logistic regression analysis showed that fabrication/falsification was a statistically significantly ($p \leq 0.007$) less frequent reason for retraction in 2011–2018 than in the reference period (1970–2000) (online supplementary information−3). Conversely, plagiarism and duplication were statistically significantly (p between 0.024 and <0.001) more frequent reasons for retraction during 2006–2018 and 2001–2018 than in 1970–2000, respectively. Temporal trends analysis showed that when fabrication/falsification or plagiarism were involved the time for retraction was statistically significantly ($p < 0.001$) shorter in 2006–2018 and 2001–2018, respectively, than in the reference period (1970–2000), with shorter times as the 21st century progressed.

Retracted medical genetics and non−medical genetics articles (online Supplementary information-4)

Table 2A (online supplementary information−5) shows the five reasons for retraction and investigation of all retracted MG and NMG articles and the number of cases that originated in the seven countries of interest, divided into six time periods. Between 5% (NMG) and 7% (MG) of retraction notices provided limited or no information. Research misconduct (ie, fabrication/falsification, plagiarism) was the reason for retraction in 28.9% and 36.5% of MG and NMG articles, respectively. Only one (out of

Table 1 Retracted genetics articles (N=1582). Reasons for retraction* and investigation, from 1970 to 2018

Reasons for retraction	n (%) (95% CI)	Reasons for retraction†	n (%) (95% CI)
Fabrication/falsification‡	377 (23.8) (21.7 to 26.0)	Results not reproducible	117 (7.4) (6.2 to 8.8)
Plagiarism§	148 (9.4) (8.0 to 10.9)	Contamination¶	27 (1.7) (1.1 to 2.5)
Duplication**	384 (24.3) (22.2 to 26.5)	Conflicts of interest	13 (0.8) (0.4 to 1.4)
Unreliable information††	196 (12.4) (10.8 to 14.1)	Fake peer review	182 (11.5) (10.0 to 13.2)
Authorship issues‡‡	159 (10.1) (8.6 to 11.6)	Error by journal/publisher	33 (20.9) (1.4 to 2.9)
Investigation§§	587 (37.1) (34.7 to 39.5)	Limited or no information	90 (5.7) (4.6 to 6.9)

The reasons for retraction and investigation of retracted articles as per the Retraction Watch database's nomenclature.

*Description of all reasons for retraction are provided in online supplementary information−2.

†Reasons with <50 retracted articles in one or both groups (medical genetics, non-medical genetics).

‡Falsification/fabrication of data, images and results, and manipulation of results and images.

§Euphemism for plagiarism and plagiarism of article, data, images or text.

¶Contamination of cell lines/tissues, materials or reagents.

**Duplication of article, data, image and text and euphemism for duplication.

††Unreliable data, image or results.

‡‡Breach of policy by author, concerns/issues about authorship, forged authorship, lack of approval from author, and objections by author.

§§Investigation by company/institution, journal/publisher, US Office of Research Integrity or by third party.

N, number of retracted articles; n, number of retracted articles because of the specified reason.

199) MG and three (out of 326) NMG retracted articles had both fabrication/falsification and plagiarism issues; thus, these two reasons for retraction were (almost) mutually exclusive. Fabrication/falsification was statistically significantly more frequently involved in NMG (27.9%) than in MG (18.6%) retracted articles. Conversely, MG retracted articles (45.2%) were statistically significantly more frequently investigated than NMG retracted reports (30.8%). There were statistically significantly more investigations conducted on retracted MG than on NMG articles in 2011−2015. China led the ranking in MG and the USA in NMG articles (table 2B). supplementary information−4

Time to retraction when research misconduct was involved is shown in table 3. With fabrication/falsification, time to retraction was statistically significantly (p=0.001) longer in NMG (mean 6.4; median 5.4 years) than in MG (mean 4.7; median 4.0 years) retracted articles for the whole study period (1970–2018). The time needed for retraction when plagiarism was involved was very similar between MG and NMG—both with means of 2.3 and medians of 1.3 years—and no statistically significant differences were found in any time period assessed except in 2016–2018. The USA was the leading country for authors of both retracted MG (n=133) and NMG (n=59) articles for fabrication/falsification, with China leading retracted MG (n=36) and NMG (n=34) articles due to plagiarism. There were several authors with a high number of retractions that affected the overall number of retractions in specific periods of time. Thus, Fazlul H Sarkar with 11 retractions and Jin Q Cheng with eight retractions represented 22% of all 86 retractions in NMG in 2006−2010; in this period Naoki Mori with eight retractions represented 15% of all retracted MG papers.

Of 207 investigations conducted by ORI on articles published during the whole study period, 62 (30%) were on genetics, with a similar percentage of retracted MG (n=29; 4.2%, 29/690) and NMG (n=33; 3.7%, 33/892) articles; the time to retraction (the time elapsed between the publication of the original article and that of the retraction notice) was not statistically significantly different (means of 4.6 and 5.1 years, medians 2.7 and 5.0, for MG and NMG, respectively). Fabrication/falsification was involved in all ORI investigated retracted articles, except in one NMG article that had unreliable results. Company/institution investigation was conducted in 17 (58.6%) of MG and 27 (81.8%) of NMG articles that were also subject to ORI investigations. Fabrication/falsification of data, data and images,

and images were involved in 30 (48.4%), 17 (27.4%), and nine (14.5%) retracted articles, respectively. Duplication of images was also involved in eight retracted articles (four MG, four NMG). All 62 retracted articles were by authors from the USA; only three MG and eight NMG articles had authors from six other countries.

Finally, table 4 shows the seven cases that were subject to civil or criminal proceedings in four different countries. Time elapsed from publication of the article to publication of the retraction notice ranged from 1.7 to 13.6 years.

DISCUSSION

The RWdb listed 1582 retracted papers on genetics. Between 1996 and 2017 there were some 975 000 worldwide genetics articles published[6]; in this same period, the RWdb listed 1476 retracted papers on genetics. Hence, we can estimate that the retraction rate for genetics articles was around 0.15%, almost four times higher than the current rate found on the RWdb for all disciplines (0.04%)[5] and almost eight times higher than that found for PubMed articles (0.02%).[1] Therefore, we should assume that genetics is a discipline with a high retraction rate.

Research misconduct was involved in 33% of all retracted genetics articles, which compares favourably with 43% for oncology,[7] 46% for all retracted articles in 2013–2015 hosted on the RWdb[8] and 49% for retracted articles in open access journals,[9] but not with 26% among BMC journals.[10] In genetics, this study showed that fabrication/falsification was much more frequent than plagiarism (24% vs 9%), a situation also found in cancer (28% vs 14%),[7] but the opposite of what happens with retracted articles in open access journals.[9 10] It is important to highlight that fabrication/falsification and plagiarisms are almost mutually exclusive: in only 0.8% (4/525) of all retracted articles were both these reasons present. Duplication, the single most prevalent reason involved in retracted genetics papers (24%), is less commonly found in oncology (18%)[7] or retracted articles in open access journals (12–16%).[9 10] It is remarkable that almost four out of 10 of all retracted genetics articles were subject to investigation. As regards to the country of origin of the authors, the seven countries assessed in our study also produced the majority of retracted articles in other analyses, although Italy, Taiwan and Iran were in the list of countries with the highest number of retractions in other analyses.[7−9]

Table 2 Retracted medical genetics and non-medical genetics articles in the Retraction Watch database, from 1970 to 2018 and in five time periods. (A) Reasons for retraction and investigation. (B) Country of origin of authors of retracted articles. Information pertaining to the seven top countries with the highest number of biomedical retracted articles[1]

Reason for retraction	1970–2018		1970–2000		2001–2005		2006–2010		2011–2015		2016–2018	
	Medical genetics N=690 n (%) (95% CI)	Non-medical genetics N=892 n (%) (95% CI)	Medical genetics N=55 n (%) (95% CI)	Non-medical genetics N=95 n (%) (95% CI)	Medical genetics N=91 n (%) (95% CI)	Non-medical genetics N=122 n (%) (95% CI)	Medical genetics N=159 n (%) (95% CI)	Non-medical genetics N=255 n (%) (95% CI)	Medical genetics N=326 n (%) (95% CI)	Non-medical genetics N=301 n (%) (95% CI)	Medical genetics N=59 n (%) (95% CI)	Non-medical genetics N=119 n (%) (95% CI)
A. Reason for retraction and investigation of retracted articles												
Fabrication/ falsification*	128 (18.6)† (15.7 to 21.7)	249 (27.9)† (25.0 to 31.0)	17 (30.9) (19.1 to 44.8)	28 (29.5) (20.6 to 39.7)	19 (20.9) (13.1 to 30.7)	47 (38.5) (29.9 to 47.8)	55 (34.6) (27.2 to 42.5)	86 (33.7) (27.9 to 39.9)	33 (10.1)† (7.1 to 13.9)	83 (27.6)† (22.6 to 33.0)	4 (6.8) (1.9 to 16.5)	5 (4.2) (1.4 to 9.5)
Plagiarism‡	71 (10.3) (8.1 to 12.8)	77 (8.6) (6.9 to 10.7)	2 (3.6) (0.4 to 12.5)	2 (2.1) (0.3 to 7.4)	8 (8.8) (3.9 to 16.6)	4 (3.3) (0.9 to 8.2)	12 (7.5) (4.0 to 12.8)	23 (9.0) (5.8 to 13.2)	36 (11.0) (7.9 to 15.0)	34 (11.3) (8.0 to 15.4)	13 (22.0) (12.3 to 34.7)	14 (11.8) (6.6 to 19.0)
Duplication§	149 (21.6) (18.6 to 24.9)	235 (26.3) (23.5 to 29.4)	0 (0.0) (0.0 to 6.5)	10 (10.5) (5.2 to 18.5)	33 (36.3) (26.4 to 47.0)	33 (27.0) (19.4 to 35.8)	46 (28.9) (22.0 to 36.6)	85 (33.3) (27.6 to 39.5)	60 (18.4) (14.3 to 23.0)	78 (25.9) (21.1 to 31.3)	10 (16.9) (8.4 to 29.0)	29 (24.4) (17.0 to 33.1)
Unreliable information¶	69 (10.0) (7.9 to 12.5)	127 (14.2) (12.0 to 16.7)	10 (18.2) (9.1 to 30.9)	15 (15.8) (9.1 to 24.7)	14 (15.4) (8.7 to 24.5)	15 (12.3) (7.0 to 19.5)	19 (11.9) (7.4 to 18.0)	37 (14.5) (10.4 to 19.4)	18 (5.5) (3.3 to 8.6)	18 (6.0) (3.6 to 9.3)	8 (13.6) (6.0 to 25.0)	16 (13.4) (7.9 to 20.9)
Authorship issues**	50 (7.2) (5.4 to 9.4)	109 (12.2) (10.1 to 14.6)	8 (14.5) (6.5 to 26.7)	12 (12.6) (6.7 to 21.0)	7 (7.7) (3.1 to 15.2)	18 (14.8) (9.0 to 22.3)	10 (6.3) (3.1 to 11.3)	34 (13.3) (9.4 to 18.1)	27 (8.3) (5.5 to 11.8)	34 (11.3) (8.0 to 15.4)	11 (18.6) (9.7 to 30.9)	8 (6.7) (2.9 to 12.8)
Investigation††	312 (45.2)† (41.5 to 49.0)	275 (30.8)† (27.8 to 34.0)	16 (29.1) (17.6 to 42.9)	25 (26.3) (17.8 to 36.4)	23 (25.3) (16.7 to 35.5)	50 (41.0) (32.2 to 50.3)	71 (44.7) (36.8 to 52.7)	86 (33.7) (27.9 to 39.9)	182 (55.8)† (50.3 to 61.3)	97 (32.2)† (27.0 to 37.8)	20 (33.9) (22.1 to 47.4)	17 (14.3) (8.5 to 21.9)
B. Country of origin of authors of retracted articles												
China	296	213	0	1	10	4	32	49	222	104	32	55
Germany	35	40	10	8	11	7	10	11	4	12	0	2
India	13	61	1	1	1	3	5	16	3	27	3	14
Japan	52	74	4	11	16	24	24	20	6	14	2	5
South Korea	27	37	1	2	11	6	6	8	8	17	1	4
UK	17	52	1	10	4	12	7	18	4	11	1	1
USA	197	329	30	51	36	70	64	107	57	79	10	22

n, number of retracted articled because of the specified reason (A) or number of retracted articles with at least one author working in the specified country (B). Many articles were authored by individuals working in more than one of the specified countries. Most articles were retracted because of two or more reasons.

The reasons for retraction and investigation of retracted articles as per the Retraction Watch database's nomenclature:

Non-medical genetics papers comprises genetic articles with many other disciplines such as anthropology, biology, biochemistry, botany, ecology, food science, microbiology, veterinary science and zoology.

*Falsification/fabrication of data, images and results, and manipulation of results and images.

†p<0.001.

‡Euphemism for plagiarism and plagiarism of article, data, images or text.

§Duplication of article, data, image and text and euphemism for duplication.

¶Unreliable data, image or results.

**Breach of policy by author, concerns/issues about authorship, forged authorship, lack of approval from author, and objections by author.

††Investigation by company/institution, journal/publisher, US Office of Research Integrity or by third party.

N, number of retracted articles.

Table 3 Retracted medical genetics and non-medical genetics articles from 1970 to 2018 and in five time periods. (A) Time (years) to retraction when scientific misconduct (ie, fabrication/falsification, plagiarism) was involved. (B) Country of origin of authors of retracted articles. Information pertaining to the seven top countries with the highest number of biomedical retracted articles[1]

Reason for retraction	1970–2018 Medical genetics N/mean time (95% CI) – median (Q1, Q3)	1970–2018 Non-medical genetics N/mean time (95% CI) – median (Q1, Q3)	1970–2000 Medical genetics N/mean time (95% CI) – median (Q1, Q3)	1970–2000 Non-medical genetics N/mean time (95% CI) – median (Q1, Q3)	2001–2005 Medical genetics N/mean time (95% CI) – median (Q1, Q3)	2001–2005 Non-medical genetics N/mean time (95% CI) – median (Q1, Q3)	2006–2010 Medical genetics N/mean time (95% CI) – median (Q1, Q3)	2006–2010 Non-medical genetics N/mean time (95% CI) – median (Q1, Q3)	2011–2015 Medical genetics N/mean time (95% CI) – median (Q1, Q3)	2011–2015 Non-medical genetics N/mean time (95% CI) – median (Q1, Q3)	2016–2018 Medical genetics N/mean time (95% CI) – median (Q1, Q3)	2016–2018 Non-medical genetics N/mean time (95% CI) – median (Q1, Q3)
A. Time to retraction due to scientific misconduct reasons												
Fabrication or falsification*	128/4.7 (4.0 to 5.3) – 4.0 (1.8, 6.5)	248/6.4 (5.8 to 7.0) – 5.4 (2.3, 9.0)	17†/7.3 (6.6 to 8.0) – 6.6 (2.7, 10.3)	28‡/10.5 (9.8 to 11.1) – 8.6 (4.4, 15.3)	19/5.9 (5.4 to 6.4) – 5.9 (1.3, 9.9)	47§/9.5 (9.3 to 9.7) – 10.3 (6.0, 13.5)	55¶**/4.8 (4.7 to 4.8) – 5.0 (2.7, 6.8)	86††‡‡/6.9 (6.8 to 6.9) – 6.7 (5.0, 9.1)	33/2.9 (2.8 to 3.0) – 2.9 (1.6, 3.9)	82§§¶¶/3.1 (3.0 to 3.1) – 2.9 (1.4, 4.7)	4/0.8 (0.3 to 1.2) – 0.7 (0.4, 1.1)	5/1.1 (0.7 to 1.6) – 1.0 (0.5, 1.5)
P value	0.001		NS		0.005		<0.001		NS		NS	
Plagiarism***	71/2.3 (1.7 to 3.0) – 1.3 (0.5, 3.0)	76/2.3 (1.7 to 3.0) – 1.3 (0.5, 3.1)	2/7.0 (0.0 to 15.6) – 7.0 (6.6, 7.5)	2/14.6 (0.0 to 38.1) – 14.6 (13.3, 15.9)	8/5.5 (4.2 to 6.8) – 4.3 (3.2, 7.1)	4/3.5 (0.1 to 6.9) – 2.2 (0.4, 5.3)	12/2.8 (2.2 to 3.4) – 0.9 (0.5, 5.3)	23/2.4 (2.2 to 2.7) – 1.5 (0.6, 4.0)	36/1.9 (1.8 to 2.0) – 1.6 (0.9, 2.4)	33/1.8 (1.7 to 1.9) – 1.3 (0.5, 3.0)	13/0.5 (0.5 to 0.6) – 0.4 (0.3, 0.5)	14/1.2 (1.1 to 1.2) – 1.0 (0.8, 1.5)
P value	NS		NS		NS		NS		NS		0.001	

	1970–2018 Medical genetics nff/np	1970–2018 Non-medical genetics nff/np	1970–2000 Medical genetics nff/np	1970–2000 Non-medical genetics nff/np	2001–2005 Medical genetics nff/np	2001–2005 Non-medical genetics nff/np	2006–2010 Medical genetics nff/np	2006–2010 Non-medical genetics nff/np	2011–2015 Medical genetics nff/np	2011–2015 Non-medical genetics nff/np	2016–2018 Medical genetics nff/np	2016–2018 Non-medical genetics nff/np
B. Country of authors of retracted articles												
China	25/36	14/34	0/0	1/0	0/2	0/0	10/6	5/2	13/23	8/18	0/7	1/12
Germany	14/1	3/0	3/0	3/0	1/0	5/0	5/1	2/0	1/0	2/0	0/0	0/0
India	15/9	3/4	0/0	0/0	0/1	0/1	2/4	3/2	11/2	0/1	1/2	0/0
Japan	33/2	19/0	1/0	5/0	5/0	11¶/1	11¶/0	13/0	5/0	1/0	1/0	0/0
South Korea	12/6	4/3	0/0	0/0	2/1	4/0	1/2	1/1	1/1	7/4§§	0/0	0/0
UK	16/1	3/2	0/0	2/0	1/1	2/0	5/1	2/0	5/0	1/1	0/0	0/0
US	133/8	59/8	12†/0	13/0	9/1	9§/1	51**/4	28†††/3	21/3	38¶¶/3	1/0	1/1

One article could have authors from more than one country; authors with at least four retractions in a specific time period are mentioned as footnotes.
The reasons for retraction as per the Retraction Watch database's nomenclature:
*Falsification/fabrication of data, images and results, and manipulation of results and images.
†Four retracted articles by Mohan K Raizada (USA).
‡Four retracted articles by Alfredo Fusco (Italy).
§Eight retracted articles by Jin Q Cheng (USA).
¶Eight retracted articles by Naoki Mori (Japan).
**Four retracted articles by Savio LC Woo (USA).
††Eleven retracted articles by Fazlul H Sarkar (USA).
‡‡Eight retracted articles by Jin Q Cheng (USA).
§§Four retracted articles by Soo Kyung Kang (South Korea).
¶¶Five retracted articles by Fazlul H Sarkar (USA).
***Euphemism for plagiarism and plagiarism of article, data, images or text.
N, number of retracted articles; NS, not significant; 1Q, 3Q, first and third quartiles; nff, number of retracted articles for fabrication/falsification in the specified country; np, number of retracted articles for plagiarism in the specified country.

Table 4 Retracted medical genetics and non-medical genetics articles from 1970 to 2018. Articles that underwent civil or criminal proceedings

Type of proceeding	Group	Country/time to retraction* (years)	Type of article	Journal, year†
Civil	Medical genetics	USA/4.8	Research article	*Urology*, 2007
	Medical genetics	Spain/6.1	Research article	*Cell Cycle*, 2012
	Non-medical genetics	USA/2.2	Research article	*Science*, 2009
	Non-medical genetics	USA/4.5	Research article	*PNAS*, 2013
Criminal	Medical genetics	Canada/13.6	Research article	*Journal Clinical Investigation*, 2002
	Medical genetics	Italy/6.8	Clinical study	*Journal Clinical Investigation*, 2007
	Non-medical genetics	Italy/1.7	Research article	*Cell Death & Disease*, 2012

*Time elapsed from publication of the article to publication of the retraction notice.
†Year of publication of the article.

This study showed that among the reasons for retraction involving research misconduct, fabrication/falsification was significantly reduced in 2011–2018 with respect to the reference period (1970–2000); conversely, plagiarism was a significantly more frequent reason for retraction in 2006–2018. Time to retraction when research misconduct reasons were involved was statistically significantly shorter in 2006–2018—which could reflect, at least in part, the higher focus of the scientific community on research integrity in recent years.

Duplication was also a statistically significantly more frequent reason for retraction during this century, than in 1970–2000. This could underlie the fact that although image editing software facilitates manipulation of images by authors, it also eases the task of editorial teams to detect it.[11] Digital manipulation of images, however, has very different degrees of severity, the most serious being the deliberate alteration of the interpretation of data.[12] Duplication is of such unprecedented magnitude in all scientific disciplines that it has been estimated some 35 000 articles might be candidates for retraction because of image duplication.[13]

The comparison between retracted MG and NMG articles showed that of the five reasons for retraction assessed, NMG had statistically significantly higher percentages of fabrication/falsification than MG retracted articles. A highly statistically significant difference was found between the percentages of retracted articles that were subject to investigation (by company/institution, journal, third party or by the ORI) in MG (45%) and NMG (31%) articles—this was due to the statistically significant differences observed in the 2011–2018 period. This strongly suggests that alleged misbehaviour in MG articles attracted more interest and was more stringently assessed than in NMG articles, perhaps reflecting the higher societal relevance of MG rather than NMG articles. The difference observed in the percentages of retracted articles that were investigated was due to the higher interest of institutions/journals/third parties—which likely were more sensitive to research integrity issues in MG articles—rather than the ORI, since this latter body conducted a similar percentage of investigations (4.2% in MG, 3.7% in NMG). However, to put these figures into perspective, we must highlight the fact that the ORI may accept the institution's findings, ask for additional evidence or conduct its own investigation.[14]

Temporal trends of retracted MG and NMG articles throughout the study period showed that time from publication to article retraction when fabrication/falsification was involved was statistically significantly longer for NMG (mean 6.4 years) than MG (mean 4.7 years) articles. This was statistically significant only for the period 2001–2010. There were no differences when plagiarism was involved (mean of 2.3 years for both MG and NMG articles).

This is the first study on retraction notices for genetics articles. The strengths and limitations of this analysis have the same source—the data contained in the RWdb. The RWdb holds the largest number of retracted articles.[5] This is a strength, since it has been shown that both PubMed and Web of Science—commonly used for studies on retractions—are inconsistent in regard to how retractions and retracted articles are labelled,[15] and therefore many could be lost for analysis. In addition, the RWdb uses a common taxonomy, which is of invaluable help when dealing with retraction notices from various editors/authors published with different formats, clarity and level of detail (reason for retraction such as 'euphemism for duplication' or 'euphemism for plagiarism' help to appropriately classify certain cases). On the other hand, the RWdb lists only a (very large) subset of all retracted articles, and therefore some retractions may have been missed. Furthermore, a number of articles could have been more precisely classified. In any case, we should acknowledge that misclassification of papers is common even at journal level.[16] In addition, and to put our findings into perspective, it should be highlighted that not all articles that should be retracted are reported as such with a retraction notice by the journal. A recent study has shown that among authors of 200 articles who were found guilty of research misconduct after ORI investigations, 41 (20.5%) articles did not have a retraction notice published.[17]

This study shows that genetics is a discipline with a relatively high retraction rate. Journal editors should have a proactive attitude toward maintaining the integrity of the published record. In addition to the available online and offline tools for detecting plagiarism, new tools are being developed to help editors to discover certain types of research misbehaviour. One of them is automated software to detect duplications, whose authors estimate that 0.6% of papers contain fraudulent images.[18] Another is a program (the 'Seek & Blastn' tool) that spots incorrect nucleotide sequence reagents reported in articles that could invalidate the results.[19] This system also allows the detection of errors unrelated to misconduct that, however, could have an important negative impact on the reliability of genetics results. As it has been recently validated,[20] it could be incorporated into the editorial process and help editors in their article screening process which could lead to retractions—as has already happened in 17 cases.[21] Although few editors currently do this with image data,[22 23] these new programs should also be used in the manuscript editorial process to prevent the publication of articles with duplicated images or mismatched gene sequences. Implementing processes to prevent publication of flawed research is always better than retracting it years later.

Correction notice The article has been corrected since it was published Online First. Fazlul Sarkar's name has been corrected in table 3 and in the text of the article.

ORCID iD
Rafael Dal-Ré http://orcid.org/0000-0002-0980-2486

REFERENCES

1 Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A* 2012;109:17028–33.
2 US Department of Health and Human Services. 42 cfr parts 50 and 93. public health service policies on research misconduct; final rule. *Fed Regist* 2005;70:28386–400.
3 Collier R. Shedding light on retractions. *CMAJ* 2011;183:E385–6.
4 Retraction Watch Database. The Retraction Watch Database, version: 1.0.5.5. Available: http://retractiondatabase.org/RetractionSearch.aspx
5 Brainard J, You J. Rethinking retractions. *Science* 2018;362:390–3.
6 . Scimago Journal and country RANK. Available: https://www.scimagojr.com/countryrank.php?category=1311&area=1300&min=0&min_type=it
7 Bozzo A, Bali K, Evaniew N, Ghert M. Retractions in cancer research: a systematic survey. *Res Integr Peer Rev* 2017;2:5.
8 Ribeiro MD, Vasconcelos SMR. Retractions covered by Retraction Watch in the 2013–2015 period: prevalence for the most productive countries. *Scientometrics* 2018;114:719–34.
9 Wang T, Xing Q-R, Wang H, Chen W. Retracted publications in the biomedical literature from open access journals. *Sci Eng Ethics* 2019;25:855–68.
10 Moylan EC, Kowalczuk MK. Why articles are retracted: a retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open* 2016;6:e012047.
11 White C. Software makes it easier for journals to spot image manipulation. *BMJ* 2007;334:607
12 Rossner M, Yamada KM. What's in a picture? The temptation of image manipulation. *J Cell Biol* 2004;166:11–15.
13 Bik EM, Fang FC, Kullas AL, Davis RJ, Casadevall A. Analysis and correction of inappropriate image duplication: the *Molecular and Cellular Biology* Experience. *Mol Cell Biol* 2018;38:pii: e00309–18.
14 Resnik DB. Research Integrity. In: *The ethics of research with human subjects*. 74. International Library of Ethics, Law, and the New Medicine. Springer, Cham, Switzerland, 2018.
15 Schmidt M. An analysis of the validity of retraction annotation in PubMed and the Web of Science. *J Assoc Inf Sci Technol* 2018;69:318–28.
16 Shu F, Julien C-A, Zhang L, Qiu J, Zhang J, Larivière V. Comparing journal and paper level classifications of science. *J Informetr* 2019;13:202–209.
17 Drimer–Batca D, Iaccarino JM, Fine A. Status of retraction notices for biomedical publications associated with research misconduct. *Res Ethics* 2019;15:1–5.
18 Acuna DE, Brookes PS, Kording KP. Bioscience–scale automated detection of figure element reuse. *bioRxiv* 2018 February 22.
19 Byrne JA, Labbé C. Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines. *Scientometrics* 2017;110:1471–93.
20 Labbé C, Grima N, Gautier T, Favier B, Byrne JA. Semi-automated fact-checking of nucleotide sequence reagents in biomedical research publications: The Seek & Blastn tool. *Plos One* 2019;14:e0213266.
21 Byrne J. We need to talk about systematic fraud. *Nature* 2019;566:9.
22 Journal of Cell Biology. Editorial policies. Data integrity and plagiarism. Available: http://jcb.rupress.org/editorial-policies#prior
23 EMBO Press. Author guidelines. Data integrity of images. Available: http://emboj.embopress.org/authorguide#scientificintegrity