



HAL
open science

The thin ret(raction) line: biomedical journal responses to incorrect non-targeting nucleotide sequence reagents in human gene knockdown publications

Jennifer A. Byrne, Yasunori Park, Rachael A. West, Amanda Capes-Davis, Bertrand Favier, Guillaume Cabanac, Cyril Labbé

► To cite this version:

Jennifer A. Byrne, Yasunori Park, Rachael A. West, Amanda Capes-Davis, Bertrand Favier, et al.. The thin ret(raction) line: biomedical journal responses to incorrect non-targeting nucleotide sequence reagents in human gene knockdown publications. *Scientometrics*, 2021, 126 (4), pp.3513-3534. 10.1007/s11192-021-03871-9 . hal-03155496

HAL Id: hal-03155496

<https://hal.science/hal-03155496>

Submitted on 7 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



The thin ret(raction) line: biomedical journal responses to incorrect non-targeting nucleotide sequence reagents in human gene knockdown publications

Jennifer A. Byrne^{1,2} · Yasunori Park² · Rachael A. West^{2,3} ·
Amanda Capes-Davis^{2,4} · Bertrand Favier⁵ · Guillaume Cabanac⁶ ·
Cyril Labbé⁷

Received: 25 September 2020 / Accepted: 15 January 2021 / Published online: 28 February 2021
© The Author(s) 2021

Abstract

The capacity of the scientific literature to self-correct is of vital importance, but few studies have compared post-publication journal responses to specific error types. We have compared journal responses to a specific reagent error in 31 human gene knockdown publications, namely a non-targeting or negative control nucleotide sequence that is instead predicted to target a human gene. The 31 papers published by 13 biomedical journals generated 26 published responses (14 retractions, 5 expressions of concern, 7 author corrections which included one resolved expression of concern) as well as 6 stated decisions to take no action. Variations in published responses were noted both between journals and by 4 journals that published different responses to at least 2 papers. A subset of published responses revealed conflicting explanations for the wrongly identified control reagent, despite 30/31 papers obtaining their gene knockdown reagents from the same external supplier. Viewed collectively, different journal responses to human gene knockdown publications with a common reagent error type suggest that editorial staff require more support to interpret post-publication notifications of incorrect nucleotide sequence reagents. We propose a draft template to facilitate the communication, interpretation and investigation of published errors, including errors affecting research reagents.

Keywords Correction · Experimental reagent · Expression of concern · Gene knockdown · Nucleotide sequence · Retraction

✉ Jennifer A. Byrne
jennifer.byrne@health.nsw.gov.au

Extended author information available on the last page of the article

Introduction

The capacity of the research literature to self-correct is of vital importance, particularly given reports of high rates of published errors within the research literature (Allison et al. 2016; Bik et al. 2016; Bik et al. 2018; Brown et al. 2018; Georgescu and Wren 2018; Guttinger and Love 2019; Nuijten et al. 2016). Errors that escape detection during both manuscript preparation and peer review can be subsequently flagged and/or corrected by approaches including retractions, expressions of concern and author corrections (Fanelli et al. 2018; Vaught et al. 2017). Despite the range of approaches that are available to correct errors within the literature, researchers have described difficulties in navigating these processes (Allison et al. 2016; Grey et al. 2020a; Malički et al. 2019; Saiz et al. 2018; Vorland et al. 2020). Post-publication review processes have been described as temporally slow, complex and time consuming (Grey et al. 2020a; Malički et al. 2019; Saiz et al. 2018), and even resulting in financial burdens to the notifying party (Allison et al. 2016). As these factors are likely to discourage researchers from communicating published errors, it is perhaps not surprising that few teams have described their efforts to correct published errors (Allison et al. 2016; Grey et al. 2020a; Malički et al. 2019; Saiz et al. 2018). More such descriptions would allow the detailed examination of post-publication review processes, which could lead to process improvements, and more timely responses to incorrect published information (Grey et al. 2020a).

We have an opportunity to compare journal responses to post-publication notifications of very similar errors, as members of our team have been describing concerns about biomedical research papers with a restricted range of error types to their corresponding journals since 2015. While only a minority of these notifications have been resolved, comparing even a limited number of journal responses can provide much needed information about how biomedical journals respond to very similar published errors.

Published nucleotide sequence reagent errors

The publications of concern that we have described to journals commonly analyse human gene function in cancer cell lines (Byrne and Labbé 2017; Labbé et al. 2019). Experiments that analyse gene function often employ short pieces of DNA or RNA that correspond to genes of interest, and these DNA or RNA sequences can be written into publications to signal exactly how specific genes were investigated. Because written DNA or RNA sequences lack obvious meaning, even short nucleotide sequences are susceptible to being transcribed incorrectly (Byrne et al. 2019; Labbé et al. 2019). Nucleotide sequences can incorporate the equivalent of spelling mistakes (Chiarella et al. 2015; Habbal et al. 2005; Labbé et al. 2019; Shannon et al. 2008) and intended nucleotide sequences can also be substituted by nucleotide sequences with entirely different identities and meanings (Byrne and Labbé 2017; Chiarella et al. 2015; Katavetin et al. 2005; Kocemba et al. 2016; Labbé et al. 2019; Tamm 2016). Whereas typographic errors within nucleotide sequence reagents can have variable significance, wrongly identified nucleotide sequence reagents will almost certainly invalidate any experiment that uses such reagents (Byrne et al. 2019; Labbé et al. 2019). Wrongly identified nucleotide sequence reagents, particularly those that recur across multiple papers, also risk being incorporated into future studies, potentially leading to failed experiments and/or publication of further incorrect data (Byrne et al. 2019; Labbé et al. 2019). Our previous and ongoing analysis of published nucleotide sequence errors

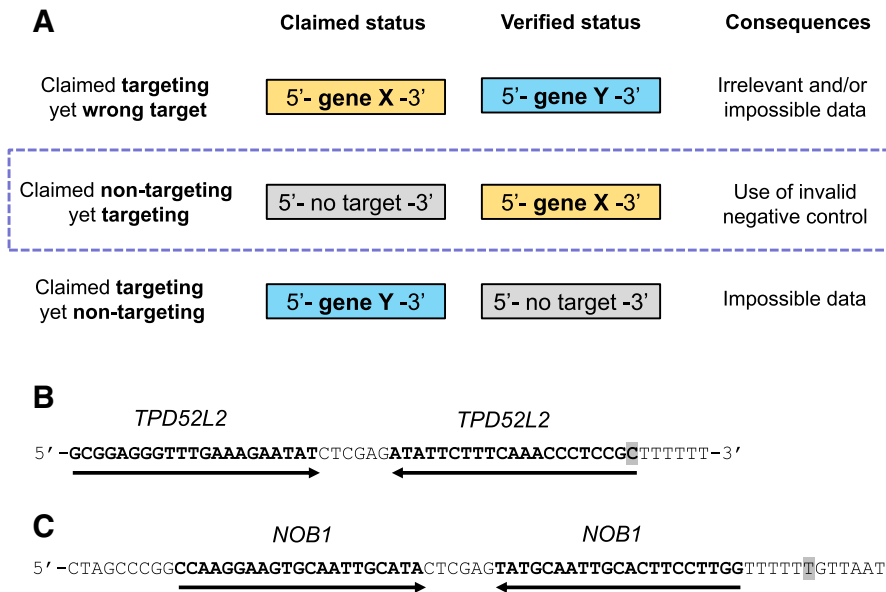


Fig. 1 a Diagrammatic representation of three different nucleotide sequence identity error types (shown at left) with claimed versus verified reagent status (centre panels) and the experimental consequences (shown at right). The error type where a claimed “non-targeting” reagent is instead predicted to target a gene is highlighted (middle row). **b, c** The incorrect non-targeting reagents Sequence A (B) and Sequence D (C). Nucleotide sequences are shown 5’-3’, with 21 nucleotide sequences that are 100% identical to the human genes *TPD52L2* (B) or *NOB1* (C) shown in bold, with arrows indicating the direction of nucleotide sequence complementarity. The presence of two inverted or self-complementary sequences is typical of shRNA reagents (Lambeth and Smith 2013). Single nucleotides that are deleted in variant sequences are highlighted in grey

has therefore focused on the detection and description of wrongly identified nucleotide sequence reagents (Byrne and Labbé 2017; Labbé et al. 2019).

We have identified and described three types of wrongly identified nucleotide sequence reagents (Byrne and Labbé 2017; Labbé et al. 2019) (Fig. 1a). The first error type represents the DNA or RNA equivalent of substituting one recognizable word for an unrelated word (Fig. 1a). Such incorrect DNA or RNA reagents will target a different gene or sequence from that intended (Byrne and Labbé 2017; Labbé et al. 2019). The second error type substitutes the DNA or RNA equivalent of a recognized word with a meaningless string of letters (Fig. 1a). Instead of targeting the stated gene or sequence, such reagents appear to not target any sequence in the species under study (Labbé et al. 2019). The third error type represents the converse problem, where a deliberately meaningless string of letters is replaced with the DNA or RNA equivalent of a recognized word (Byrne and Labbé 2017; Labbé et al. 2019) (Fig. 1a). Whereas non-targeting reagents should not target any gene or sequence in the species under study, incorrect “non-targeting” reagents show unexpected homology to a specific gene (Fig. 1b, c) (Byrne and Labbé 2017; Labbé et al. 2019).

Incorrect non-targeting nucleotide sequence reagents

To compare how journals respond to wrongly identified nucleotide sequence reagents, we will describe journal responses to descriptions of incorrect non-targeting reagents (Byrne and Labbé 2017; Labbé et al. 2019). We have focused upon incorrect non-targeting reagents for several reasons. Firstly, the description of a single incorrect non-targeting reagent typically represents a fatal flaw in any gene knockdown publication, where every individual gene knockdown is typically paired with the same non-targeting control (Han 2018). The description of an incorrect non-targeting control is therefore a very serious error that is sufficient to invalidate all affected gene knockdown experiments. Secondly, incorrect non-targeting control reagents represent the most frequent nucleotide sequence identity error type that we initially reported within a cohort of 48 strikingly similar human gene knockdown papers (Byrne and Labbé 2017). Our subsequent notifications of these and other papers with incorrect non-targeting controls mean that journal responses are more likely to concern papers with this error type. Thirdly, subsequent analyses suggest that incorrect non-targeting reagents represent a robustly detected error type (Labbé et al. 2020), as such reagents have “gained” a gene targeting function that is unambiguous and therefore easy to detect (Fig. 1b, c). Finally, as the same non-targeting reagent can be legitimately applied for the analysis of many different genes, incorrect non-targeting controls can be found across many papers (Byrne and Labbé 2017; Labbé et al. 2019). This means that by searching for a small number of incorrect control reagents, we can identify a larger number of fatally flawed gene knockdown papers (Byrne and Labbé 2017). By then describing such errors to journals, we can eventually compare journal responses to a set of very similar papers that describe the application of the same gene knockdown technique, and even the same incorrect negative control.

Our first report of incorrect nucleotide sequence reagents in gene knockdown papers identified two incorrect non-targeting reagents in multiple papers (Byrne and Labbé 2017) (Fig. 1b, c). These reagents named Sequence A or Sequence D are indicated to represent *TPD52L2* or *NOBI* targeting short hairpin RNA's (or shRNA's) as these reagents show inverted or self-complementary sequences that are identical to *TPD52L2* (Sequence A) or *NOBI* transcripts (Sequence D), respectively (Byrne and Labbé 2017) (Fig. 1b, c). While both Sequence A and Sequence D have been correctly described as *TPD52L2* and *NOBI* targeting reagents, they have also been incorrectly described as “non-targeting” controls in numerous publications (Byrne and Labbé 2017; Labbé et al. 2019). We therefore compared 32 available journal responses to 31 human gene knockdown papers that commonly examined the functions of single human genes in human cell lines corresponding to different cancer types, with each paper also specifying an incorrect non-targeting control.

Materials and methods

Identification of journal responses to publications with incorrect non-targeting control reagents

Most journal responses followed our notifications of papers that described Sequence A or Sequence D, or a very highly related variant sequence (Fig. 1b, c) (Byrne and Labbé 2017). Concerns were communicated to journals by emails written and sent by JAB and copied to CL, with follow-up communications also occurring through emails written and

sent by JAB. Emails were sent to the correspondence address listed on journal websites and were copied to one or more editors or managing editors, using either journal-supplied email addresses, and/ or email addresses that were retrieved by JAB using names and institutional addresses as Google search queries. Emails described concerns about structural and textual similarities with other human gene knockdown papers and the presence of an incorrect non-targeting control reagent (Sequence A, Sequence D, or a very highly related variant sequence) (Byrne and Labbé 2017). Where concerns involved multiple papers published by the same journal, concerns were described in single emails that described up to 6 individual papers. Errors in papers that described the analysis of the *TPD52L2* gene were initially communicated to their publishing journals in June 2015, with other papers communicated in January 2017, following the description of 48 strikingly similar human gene knockdown papers (Byrne and Labbé 2017). The communication of one paper was delayed until January 2018, after the paper was initially communicated to the wrong journal by JAB in January 2017.

Journal responses to 29 human gene knockdown publications with incorrect non-targeting controls were identified through PubMed and/or Google Scholar searches using publication titles as search queries. We also identified journal responses to two additional human gene knockdown papers that refer to the use of Sequence A as a “non-targeting” control (Table 1). One journal response resulted from the journal *BioMed Research International* identifying and investigating a human gene knockdown paper (BioMed Research International 2017a) in response to our notification of a very similar paper (BioMed Research International 2017b) (Table 1). The second additional response followed the recognition of Sequence A as an incorrect non-targeting control by the authors of an *OncoTargets and Therapy* publication (Retraction 2017) (Table 1). This published response was identified using Sequence A as a Google Scholar search query, as previously described by Byrne and Labbé (2017) and described in further detail below.

Analysis of journal responses

Times for journals to respond to notifications (measured in months) were calculated according to the month and year when concerns were first communicated by JAB to the publishing journal, and the month and year in which the journal response was published online, as identified through PubMed and/or Google Scholar searches of publication titles. In the case of the single journal that decided to take no action in response to our concerns, the response time was calculated from the month and year of our first email to the month and year of the journal’s email stating their final decision to take no action. In the case of the human gene knockdown paper that was identified and investigated by *BioMed Research International* (BioMed Research International 2017a), the journal’s response to this paper was considered to be temporally linked to our description of another *BioMed Research International* human gene knockdown paper (BioMed Research International 2017b) (Table 1). This was supported by the journal’s published response referring to the *BioMed Research International* paper that we communicated (Table 2), and by the two journal responses being published in the same month and year (Fig. 2). Journal time to respond was not calculated for the single response that was initiated by the authors (Retraction 2017).

All published journal responses were visually inspected to determine whether individual responses referred to the incorrect non-targeting control, and whether authors

Table 1 Summary of 31 human gene knockdown publications employing human cancer cell lines that describe an incorrect non-targeting gene knockdown reagent

| # | PMID | Journal, Publisher | Year | Gene ^a | Human cancer type | Google Scholar citations $n =$ | Incorrect non-targeting control reagent ^b | Journal response | Notes |
|----|----------|---|------|-------------------------|---------------------|--------------------------------|--|-----------------------|--|
| 1 | 25301753 | Acta Biochem Biophys Sin, Oxford Academic | 2014 | <i>GPRI37</i> | Colon | 11 | Sequence D | Retraction | Variant Sequence D ^c |
| 2 | 27044563 | | 2016 | <i>MYO6</i> | Colorectal | 10 | Sequence A | Correction | |
| 3 | 25889525 | Biol Res, BMC | 2015 | <i>USP39</i> | Liver | 15 | Sequence A | Expression of concern | |
| 4 | 27177590 | | 2016 | <i>ZFR</i> | Pancreatic | 12 | Sequence A | | |
| 5 | 26448925 | BioMed Res Int, Hindawi | 2015 | <i>lincRNA K1AA0125</i> | Gallbladder | 16 | Sequence A | Retraction | |
| 6 | 27429979 | | 2016 | <i>PTGRI</i> | Non-small cell lung | 7 | Sequence A | | Paper identified by journal |
| 7 | 25322973 | Biotech App Biotech, Wiley | 2015 | <i>PP5</i> | Colorectal | 10 | Sequence A | No action | |
| 8 | 25471990 | | 2015 | <i>GPRI37</i> | Pancreatic | 12 | Sequence D | | |
| 9 | 25496438 | | 2015 | <i>GPRI37</i> | Bladder | 5 | Sequence A | | |
| 10 | 25524330 | | 2015 | <i>GPRI37</i> | Medullo-blastoma | 7 | Sequence A | | |
| 11 | 26601916 | | 2017 | <i>linc-ITGB1</i> | Breast | 16 | Sequence A | | Reagents supplied by Genepharma Co. (Shanghai) |
| 12 | 27174009 | | 2017 | <i>MPP8</i> | Colon | 6 | Sequence A | | |
| 13 | 25629696 | Cancer Biother | 2015 | <i>TPD52L2</i> | Breast | 9 | Sequence A | Retraction | |
| 14 | 25746840 | Radiopharm, Mary Ann Liebert | 2015 | <i>TPD52L2</i> | Gastric | 6 | Sequence D | | |
| 15 | 25409460 | | 2014 | <i>RPS15A</i> | Osteosarcoma | 16 | Sequence A | Correction | |
| 16 | 25279970 | | 2014 | <i>ADRBK1</i> | Breast | 7 | Sequence A | | |
| 17 | 25262828 | Cell Biol Int, Wiley | 2015 | <i>TPD52L2</i> | Oral | 5 | Sequence D | Retraction | |
| 18 | 25215606 | Chem Biol Drug Design, Wiley | 2015 | <i>PDLIM5</i> | Gastric | 5 | Sequence D | Retraction | |
| 19 | 25893892 | | 2015 | <i>linc-ITGB1</i> | Gallbladder | 19 | Sequence A | | |
| 20 | 25643992 | | 2015 | <i>MYO6</i> | Lung | 9 | Sequence A | Correction | |
| 21 | 25703929 | | 2015 | <i>MYO6</i> | Hepatocellular | 8 | Sequence A | | |
| 22 | 25676706 | | 2015 | <i>TPTEP1</i> | Gallbladder | 6 | Sequence A | | |

Table 1 (continued)

| # | PMID | Journal, Publisher | Year | Gene ^a | Human cancer type | Google Scholar citations <i>n</i> = | Incorrect non-targeting control reagent ^b | Journal response | Notes |
|----|----------|---|------|-------------------|-------------------------|-------------------------------------|--|-----------------------|-----------------------------------|
| 23 | 25932170 | Int J Clin Exp Med, e-Century | 2015 | TPD52L2 | Liver | 10 | Sequence A | Retraction | Variant Sequence A ^c |
| 24 | 26550235 | | 2015 | TCTN1 | Medullo-blastoma | 6 | Sequence A | | |
| 25 | 26823768 | Int J Clin Exp Path, e-Century | 2015 | PTGRI | Gastric | 4 | Sequence A | Retraction | |
| 26 | 26472971 | J Breast Cancer, Korean Breast Cancer Society | 2015 | PP4R1 | Breast | 2 | Sequence A | Retraction | |
| 27 | 26937209 | Korean J Physiol Pharmacol, Medrang | 2016 | MYO6 | Glioma | 6 | Sequence A | Retraction | |
| 28 | 25432697 | Med Oncol, Springer | 2015 | IL1R2 | Osteosarcoma | 9 | Sequence A | Expression of concern | Updates expression of concern |
| 29 | 26008152 | | 2015 | eIF3d | Non-small cell lung | 15 | Sequence A | Correction | |
| 30 | 28652764 | Oncotargets Ther, Dovepress | 2017 | SUN1 | Lung | 3 | Sequence A | Retraction | Paper communicated by the authors |
| 31 | 26056476 | | 2015 | ICT1 | Glioblastoma multiforme | 7 | Sequence A | Expression of concern | |

Publications are ranked by journal name in alphabetical order, and then by publication year

^aGene identifiers shown in bold represent human genes that were studied in at least two of the papers listed

^bSequences shown in bold can be identified in the corresponding publication when used as a text string in a Google Scholar search

^cVariant forms of Sequence A and Sequence D as shown in Figs. 1b and 1c, respectively

Table 2 Comparison of 26 published journal responses to 25 human gene knock down papers that describe an incorrect non-targeting nucleotide sequence reagent

| # | PubMed ID | Journal | Journal response | Reagent error described in response? | Author involvement in response? | Response refers to reagent supplier? | Non-targeting sequence corrected? | Notes |
|----|-----------|---------------------------|-----------------------|--------------------------------------|---------------------------------|--------------------------------------|-----------------------------------|--|
| 1 | 29726888 | Acta Biochem Biophys Sin | Retraction | No | No | No | No | "Experimental data/plasmid vectors" |
| 2 | 29608642 | | Correction | Yes | Yes | Yes | Yes ^a | "Technical problems of the manufacturer" |
| 3 | 29933754 | Biol Res | Expression of concern | Yes | Yes | No | No | Authors agreed with expression of concern |
| 4 | 29933757 | | Expression of concern | Yes | No | No | No | Authors did not respond |
| 5 | 28929110 | BioMed Res Int | Retraction | Yes | Yes | No | No | "Nontargeting sequence plasmid is actually an empty vector" |
| 6 | 28932745 | | Retraction | Yes | No | No | No | Paper identified by journal, authors did not respond |
| 7 | 29195061 | Cancer Biother Radiopharm | Retraction | Yes | Yes | No | No | Authors indicated "copy paste error", agreed to retraction. Retraction corrected by journal 10/2018 ^c |
| 8 | 29195059 | | Retraction | Yes | Yes | No | No | Authors disagreed with concerns. Retraction corrected by journal 09/2018 ^c |
| 9 | 31329732 | | Correction | Yes | Yes | Yes | Yes ^a | Joint correction, "Holly Biotechnologies provided...wrong product specification" |
| 10 | | | | | | | | |
| 11 | 26909548 | Cell Biol Int | Retraction | No | Yes | No | No | Retraction based on incorrect cell line, authors agreed |
| 12 | 30468310 | Chem Biol Drug Design | Retraction | Yes | No | No | No | "Incorrect shRNA sequences being used" |
| 13 | 30267485 | | Retraction | Yes | Yes | No | No | "Incorrect shRNA sequences being used" |
| 14 | 30171676 | | Correction | Yes | Yes | No | Yes ^a | "Clerical error" involving control plasmid sequence |
| 15 | 30267486 | | Correction | Yes | Yes | No | Yes ^b | "Clerical error" involving control plasmid sequence |
| 16 | 30267484 | | Correction | Yes | Yes | No | No | Authors agreed that negative control is targeting but "other control is still correct" |

Table 2 (continued)

| # | PubMed ID | Journal | Journal response | Reagent error described in response? | Author involvement in response? | Response refers to reagent supplier? | Non-targeting sequence corrected? | Notes |
|----|-----------|----------------------------|-------------------------------------|--------------------------------------|---------------------------------|--------------------------------------|-----------------------------------|--|
| 17 | 29874322 | Int J Clin Exp Med | Retraction | No | Yes | No | No | "Experiments were sourced out to a biotechnology company" |
| 18 | 29874342 | | Retraction | No | No | No | No | "Experimental defects" mentioned, authors did not respond |
| 19 | 29874340 | Int J Clin Exp Path | Retraction | No | No | No | No | "Experimental defects" mentioned, authors did not respond |
| 20 | 29628991 | J Breast Cancer | Retraction | Yes | Yes | No | No | "Sequence of ... negative control lentivirus ... written wrongly" |
| 21 | 28883760 | Korean J Physiol Pharmacol | Retraction | Yes | Yes | No | No | "Inaccurate use of siRNA sequences", agreed to retraction |
| 22 | 30120568 | Med Oncol | Expression of concern | Yes | No | No | No | Authors did not respond |
| 23 | 30121714 | | Expression of concern | Yes | No | No | No | Authors did not respond |
| 24 | 31053950 | | Correction to Expression of Concern | Yes | Yes | No | Yes ^a | Correction to Expression of concern 30121714, non-targeting control sequence corrected |
| 25 | 29343968 | OncoTargets Ther | Retraction | Yes | Yes | Yes | Yes ^a | "Authors cannot confirm ... sequence carried by the control group" |
| 26 | 30464508 | | Expression of concern | Yes | No | No | No | Journal "unable to make a definitive conclusion" about incorrect non-targeting reagent |

Journal responses are listed according to journal name in alphabetical order

^a5'-TTCCTCCGAACGGTGCACGTCCTCGAGACGTGACACGTTCCGGAGAAATTTT-3' provided as corrected non-targeting control sequence

b5'-TTCCTCCGAACGGTGCACGTCCTCGAGACGTGACACGTTCCGGAGAA-3' provided as corrected non-targeting control sequence

^aRetraction initially contained incorrect information about nucleotide sequence errors, subsequently corrected through use of blastn as described by Byrne and Labbé (2017)

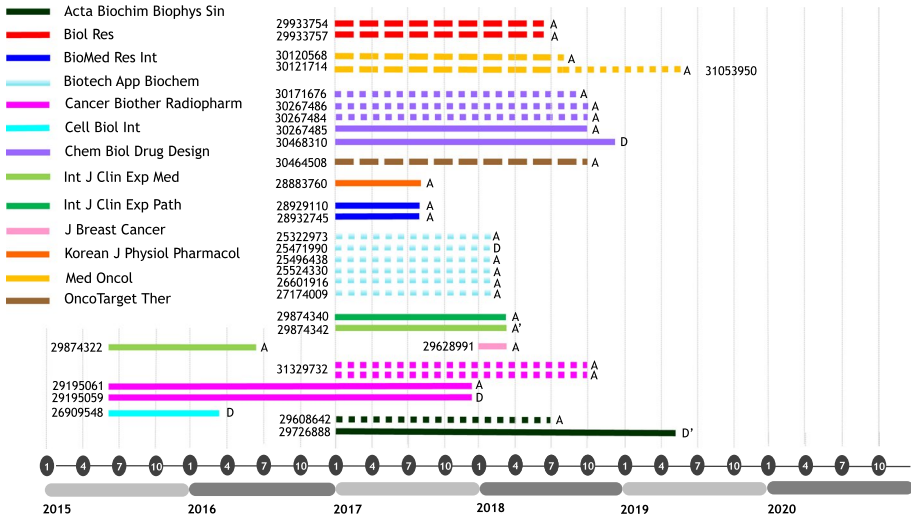


Fig. 2 Diagrammatic representation of journal response times to 30 gene knockdown papers, from communication to journal decision, according to the timeline on the X axis (years, numbers refer to the first month in each quarter). The retraction which was requested by the authors (Retraction 2017) is not shown. Each horizontal line represents a single publication, with different colours representing different journals according to the key shown at left. Published retractions, expressions of concern and author corrections are represented by solid lines, broken lines, and dotted lines, respectively. Decisions to take no action are indicated by shaded dotted lines. The incorrect non-targeting control is shown at the right of each publication (A = Sequence A, A' = variant Sequence A, D = Sequence D, D' = variant Sequence D). PubMed ID's of retractions, expressions of concern, corrections or papers for which no action was taken are shown at left, apart from the single author correction that resolved an earlier expression of concern, where the PubMed ID of the correction is shown at right. One author correction (PubMed ID 31329732) jointly corrected two papers

were consulted prior to publication of the response. As two author corrections referred to a common reagent description error made by the same reagent supply company (Correc-tion 2018), original publications (Table 1) were examined to identify which reagent supply companies were cited as having provided either expression plasmids which are likely to encode shRNA reagents and/or shRNA sequence reagents. The terms “Shanghai” and “HollyBio” were used in combination and/or individually to query pdf files for reference to Shanghai HollyBio as the supplier of gene knockdown reagents. Text references to either “Shanghai HollyBio” or “HollyBio”, where this referred to a Shanghai-based company, were accepted as corresponding to the same company. Where other suppliers were stated to have provided either expression plasmids and/or shRNA sequence reagents for gene knockdown experiments, the names and geographic locations of these suppliers were also recorded.

Identification of other human gene knockdown papers that describe Sequence A or Sequence D as a non-targeting control reagent

Google Scholar searches were performed in July 2020 to identify other human gene knock-down publications that employed either Sequence A or Sequence D as a “non-targeting”

control (Byrne and Labbé 2017). Text strings corresponding to either Sequence A (GCG GAGGGTTTCAAAGAATATCTCGAGATATTCTTTCAAACCCTCCGCTTTTTT) or Sequence D (CTAGCCCGCCAAGGAAGTGCAATTGCATACTCGAGTATGCAATTGCACTTCCTTGGTTTTTTGTAAAT) were employed as individual text queries in Google Scholar searches (Byrne and Labbé 2017). The pdf file of each paper was then inspected to confirm (i) the presence of either Sequence A or Sequence D and (ii) either “Shanghai HollyBio” or “HollyBio” as a Shanghai-based company, or any other gene knockdown reagent supplier. Papers were also screened using Seek and Blastn, with nucleotide sequence identities manually verified as previously described (Labbé et al. 2019). The PubMed ID of each identified paper was used to query email correspondence archives, to verify whether any papers had been previously communicated to their publishing journals.

Citation counts for human gene knockdown papers

Citation counts for human gene knockdown papers are those identified by Google Scholar in September, 2020. Citations within retraction notices, expressions of concern and/or author corrections to the same paper were excluded.

Results

Journal responses to human gene knockdown publications with an incorrect non-targeting control

We analysed journal responses to 31 gene knockdown papers that commonly examined single human genes in cancer cell lines corresponding to different cancer types (Table 1). The 31 papers targeted 21 different human genes, with 5 genes being represented across multiple papers and human cancer types (Table 1). All 31 papers shared the fatal error of describing a “non-targeting” control reagent that is predicted to target a human gene (Byrne and Labbé 2017) (Tables 1, 2). Most (26/31) papers specified Sequence A as their “non-targeting” control, which is predicted to target *TPD52L2* (Fig. 1b), whereas the remaining 5 papers described Sequence D, which is predicted to target *NOBI* (Fig. 1c, Table 1) (Byrne and Labbé 2017). The 31 papers were published by 13 different journals between 2014 and 2017 (Table 1). Citation numbers for individual publications ranged from 2 to 19 citations/ paper, and as of September 2020, the 31 papers had been collectively cited 279 times (Table 1).

The 31 papers generated 32 post-publication responses, which represented 14 retractions, 5 expressions of concern, 7 author corrections (including one author correction that resolved an earlier expression of concern), and 6 stated decisions to take no action (Tables 1, 2). As described in the Methods, we could measure response times for 31/32 journal responses, excluding the single author-initiated retraction (Retraction 2017) (Table 2, Fig. 2). Overall, journal response times varied from 2 to 30 months (Fig. 2). Times to retraction varied from 2 to 30 months for 13 retractions that were published from 2016 to 2019, whereas times to publication of 5 expressions of concern and 7 author corrections ranged from 17–21 and 19–28 months, respectively. Most or all author corrections and expressions of concern were published during 2018, with the single correction that resolved an expression of concern being published in 2019 (Fig. 2). One journal required

13 months to communicate their decision to take no action over 6 papers (Fig. 2). Only 6/31 responses were published within 12 months, all of which were retractions (Fig. 2).

We considered whether the 26 published responses (retractions, expressions of concern, or author corrections) referred to the “non-targeting” reagent and/ or involved input from study authors (Table 2). Most (21/26) published responses (9/14 retractions, 5/5 expressions of concern, 7/7 author corrections) referred to the incorrect control reagent (Table 2). Of the remaining 5 retractions, one referred indirectly to “experimental data/ plasmid vectors”, two referred to “experimental defects”, one referred to “experiments... sourced out to a biotechnology company”, and the final retraction referred to the use of a contaminated human cell line (Table 2). Similarly, most (17/26) published responses (9/14 retractions, 1/5 expressions of concern, 7/7 author corrections) mentioned author involvement or agreement with the published response, whereas the remaining 9 responses (5/9 retractions, 4/5 expressions of concern) either did not mention author involvement or specified that authors could not be contacted (Table 2). Overall, 15 published responses (7/14 retractions, 1/5 expressions of concern and 7/7 author corrections) referred to the wrongly identified control reagent and specified author involvement in the published response.

When responses were considered according to publishing journal, 4/13 journals responded to a single paper, publishing a retraction in each case, whereas the remaining 9 journals responded to at least two papers. Five of these 9 journals issued the same response to multiple papers (either retractions, expressions of concern, or the decision to take no action), whereas 4 journals published different responses to different papers (Fig. 2, Table 2). *Acta Biochimica Biophysica Sinica* published an author correction to one paper but later retracted a second paper (Fig. 2, Table 2). *Cancer Biotherapy and Radiopharmaceuticals* retracted two papers that were communicated in June 2015 and subsequently published author corrections of two papers that were communicated in January 2017 (Fig. 2, Table 2). *Chemical Biology and Drug Design* retracted two papers and published author corrections of 3 papers (Fig. 2, Table 2). One of these author corrections is factually incorrect, because the authors denied the significance of the incorrect non-targeting control (Table 2). Finally, *OncoTargets and Therapy* retracted one paper in response to an author description of an incorrect non-targeting reagent (Sequence A) (Table 2) and subsequently published an Expression of Concern to another paper that employed the same “non-targeting” control (Fig. 2, Table 2). Whereas the author-initiated retraction stated that the authors had been advised of an error affecting the control sequence by the supplier, the subsequently published expression of concern for a different paper indicated that *OncoTargets and Therapy* “was unable to make a definitive conclusion” about the same wrongly identified control reagent (Table 2).

External suppliers of “non-targeting” control reagents

We noted that most (6/7) author corrections replaced Sequence A with non-targeting sequences that are identical over 44 nucleotides (Table 2). A joint correction of two *Cancer Biotherapy and Radiopharmaceuticals* papers stated that the non-targeting control sequence was wrongly specified by the supplier, leading to the wrong control shRNA sequence (Sequence A) being supplied in the product specification (Correction 2018) (Table 2). Although the published correction referred to the reagent supplier as Holly Biotechnologies (Correction 2018), the corresponding *Cancer Biotherapy and Radiopharmaceuticals* papers referred to the gene knockdown reagent supplier as Shanghai HollyBio or Hollybio Shanghai (Zhang et al. 2014a, b).

Text analyses revealed that almost all (30/31) gene knockdown papers referred to Shanghai HollyBio or Hollybio based in Shanghai as the supplier of either expression plasmids and/or gene knockdown nucleotide sequence reagents. These 30 papers included all 26 papers with published responses and 5/6 papers where the journal elected to take no action, with the remaining paper naming another Shanghai-based company GenePharma as the supplier of expression plasmids and nucleotide sequence reagents (Table 1). We therefore examined all published responses for explanations for the incorrect control reagent. While no other journal responses beyond the corrections published by *Cancer Biotherapy and Radiopharmaceuticals* named any reagent supply company, the author correction published by *Acta Biochimica Biophysica Sinica* referred to “technical problems of the manufacturer” (Table 2). Five retractions confirmed the use of incorrect reagents, either by stating that the authors agreed that incorrect sequences had been used (*Chemical Biology and Drug Design*, *Korean Journal of Physiology and Pharmacology*), or that the negative control sequence was written wrongly, without further explanation (*Journal of Breast Cancer*), or that the control plasmid was in fact an empty vector (*BioMed Research International*) or that authors could not confirm that the corrected control sequence provided by the supplier had been used in experiments (*Oncotargets and Therapy*) (Table 2).

Additional human gene knockdown papers that describe Sequence A or Sequence D as “non-targeting” control reagents

We have previously used nucleotide sequences as text strings in Google Scholar searches to identify papers that describe incorrect non-targeting controls (Byrne and Labbé 2017). In addition to papers listed in Table 1, Google Scholar searches performed in July 2020 identified 19 human gene knockdown papers that described either Sequence A ($n=15$ papers) or Sequence D ($n=4$ papers) as the “non-targeting” control. Ten of these 19 papers have been previously described (Byrne and Labbé 2017; Labbé et al. 2019) and communicated to their publishing journals (Table 3). The 19 papers were published between 2015 and 2019 by 13 journals and examined 13 different human genes in cell lines corresponding to 13 human cancer types (Table 3). Over half (7/13) of these human genes were represented across multiple papers and cancer types (Tables 1, 3). Most (11/19) papers specified that gene knockdown reagents were supplied by Shanghai HollyBio, with 5 papers naming 4 other suppliers (GenePharma, GeneChem, Hanbio, ShanghaiBio) that were commonly based in Shanghai, China (Table 3). In addition to their wrongly identified non-targeting reagent, Seek and Blastn analyses supported by manual verification found that 6/19 papers also described 2–5 wrongly identified targeting reagents, most of which were predicted to target genes other than those claimed in the text (Supplementary Table 1). Citation numbers for individual publications ranged from 1 to 54 citations/ paper, and as of September 2020, the 19 papers had been collectively cited 229 times (Table 3).

Discussion

Comparisons of 32 journal responses to 31 human gene knockdown papers have highlighted the wide range of responses by 13 journals to notifications of incorrect control reagents that represent fatal errors in gene knockdown studies. Journal responses ranged from issuing retractions to deciding to take no action, with variations in responses being noted

Table 3 Additional human gene knockdown publications employing human cancer cell lines that describe an incorrect non-targeting nucleotide sequence reagent, identified using either Sequence A or Sequence D as a Google Scholar search query

| # | PMID ^a | Journal, Publisher | Year | Gene ^b | Human cancer type | Google Scholar citation $n =$ | Incorrect non-targeting reagent | Reagents supplied by Shanghai HollyBio? | Incorrect targeting reagents? ^c |
|----|-------------------|----------------------------|------|-------------------------|---------------------|-------------------------------|---------------------------------|---|--|
| 1 | 30103211 | Aging, Impact Journals | 2018 | <i>ANLN</i> | Hepatocellular | 17 | Sequence D | No | No |
| 2 | 30279202* | Biosci Rep, Portland Press | 2018 | <i>linc-ITGB1</i> | Liver (Hepatoma) | 4 | Sequence A | No (GenePharma) | Yes |
| 3 | 31216997 | BMC Cancer, BMC | 2019 | <i>TOP2A</i> | Bladder | 6 | Sequence D | No (Shanghai GenePharma, Hambio) | No |
| 4 | 29201168* | Exp Ther Med, Spandidos | 2017 | <i>linc-ITGB1</i> | Hepatocellular | 3 | Sequence A | No (Shanghai GenePharma) | Yes |
| 5 | 28203520 | FEBS Open Bio, Wiley | 2016 | <i>EIF3D</i> | Ovarian | 6 | Sequence A | Yes | No |
| 6 | 26581508** | Gene, Elsevier | 2016 | <i>HNRNP1</i> | Lung | 33 | Sequence A | Yes | No |
| 7 | 26847263 | Int J Oncol, Spandidos | 2016 | <i>RPS15A</i> | Colorectal | 22 | Sequence A | Yes | Yes |
| 8 | 25223638** | Mol Cell Biochem, Springer | 2015 | <i>ANLN</i> | Breast | 54 | Sequence D | Yes | No |
| 9 | 27175747* | Mol Med Reports, Spandidos | 2016 | <i>USP39</i> | Gastric | 6 | Sequence A | Yes | No |
| 10 | 30569143 | | 2019 | <i>RPS15A</i> | Kidney | 1 | Sequence A | No (Shanghai GeneChem Co) | No |
| 11 | 27313751*** | Oncol Lett, Spandidos | 2016 | <i>MPHOSPH8 or MPP8</i> | Medullary Thyroid | 2 | Sequence A | No (Genechem, ShanghaiBio) | No |
| 12 | 27446435* | | 2016 | <i>ZFP91-P</i> | Pancreatic | 11 | Sequence D | Yes | Yes |
| 13 | 28529588 | | 2017 | <i>CALB1</i> | Osteosarcoma | 1 | Sequence A | Yes | No |
| 14 | 27431378 | Oncol Rep, Spandidos | 2016 | <i>MYO6</i> | Prostate | 11 | Sequence A | Yes | No |
| 15 | 27109614 | | 2016 | <i>TSP50</i> | Non-small cell lung | 5 | Sequence A | No | No |
| 16 | 28260108 | | 2017 | <i>GATAD2A</i> | Thyroid | 4 | Sequence A | Yes | Yes |

Table 3 (continued)

| # | PMID ^a | Journal, Publisher | Year | Gene ^b | Human cancer type | Google scholar citation <i>n</i> = | Incorrect non-targeting reagent | Reagents supplied by Shanghai HollyBio? | Incorrect targeting reagents? ^c |
|----|-------------------|-------------------------|------|----------------------|-------------------|------------------------------------|---------------------------------|---|--|
| 17 | 25833696** | Tumor Biol, SAGE | 2015 | <u><i>RPS15A</i></u> | Lung | 18 | Sequence A | Yes | No |
| 18 | 26303214** | World J Surg Oncol, BMC | 2015 | <u><i>LSP39</i></u> | Medullary Thyroid | 15 | Sequence A | Yes | No |
| 19 | 27130037** | | 2016 | <u><i>RPS15A</i></u> | Glioblastoma | 10 | Sequence A | No | Yes |

Publications are ranked according to journal name in alphabetical order, and then by publication year

^aAsterisked PubMed ID's represent papers that have been communicated to their publishing journal and for which no response has yet been received (as of December 2020). Errors in papers with single or double asterisks were communicated in either May 2019 or January 2017, respectively

^bGene identifiers shown in bold represent human genes that were studied in more than one paper. Gene identifiers that are underlined represent human genes analysed in other gene knockdown papers that describe the “non-targeting” reagents Sequence A or Sequence D (Table 1). Gene identifiers shown underlined and in bold meet both criteria

^cIncorrect targeting nucleotide sequence reagents are listed in Supplementary Table 1

both between different journals and by 4 individual journals. Viewed individually, many published journal responses are reasonable. Retractions may have reflected numerous journals' interpretation of the fatal error represented by the incorrect control reagent, which was also described in the context of other concerns regarding textual and structural similarity with other human gene knockdown papers (Byrne and Labbé 2017). Expressions of concern may also have been considered reasonable responses while journal investigations continue (Vaught et al. 2017), and we note that most expressions of concern did not specify author involvement or stated that authors could not be contacted. Where authors responded with plausible explanations for the incorrect control sequence, it is also reasonable to provide an opportunity to correct the published record. Nonetheless, given the collective range of journal responses, we set out to explain why variable and at times inconsistent responses were issued to highly similar human gene knockdown papers with the same reagent error type.

We recognize that the present analysis was not designed as a controlled study, and although the 31 papers shared striking similarities and an incorrect non-targeting control, these papers were not identical. Some papers presented with targeting nucleotide sequence reagent errors, in addition to the incorrect non-targeting control (Byrne and Labbé 2017; Labbé et al. 2019). While the presence of additional errors may have affected journal responses, the single “non-targeting” control reagent remained sufficient to invalidate all gene knockdown results described in each paper. The range of responses suggests that while most journals recognized the described error, some journals variably interpreted its significance. Particular responses indicated uncertainty about the significance of the “non-targeting” control, such as responses that did not mention the reagent error, the publication of an author correction that dismissed the importance of an incorrect non-targeting control (Table 2), and most concerningly, the single journal that elected to take no action over 6 papers (Table 1).

We recognize that the interpretation of errors that are notified by third parties may be more complex than errors described by authors, and that nucleotide sequence reagent errors can be difficult to understand, particularly those affecting non-targeting controls. As we notified concerns about 29/31 papers, we must share responsibility for journals having an incomplete understanding of the significance of the incorrect control reagent. We described our concerns by email, and varied journal responses could suggest that emails may inadequately explain concepts such as incorrect gene knockdown controls. Our emails may have provided insufficient information and/or may have used a format that did not allow journal staff to quickly or easily understand the described errors and concerns, such as the status of an incorrect non-targeting control as a fatal error. For this reason, we suggest the derivation and adoption of structured, standardized and ideally universal templates for post-publication error communication (Table 4). Such templates would be co-designed by researchers, editorial staff and publishers to provide the information that different stakeholders need to assess published errors.

Structured and shared error communication templates would present many advantages. Such templates could be used by both authors and third parties and completed templates could then be shared with authors and peer reviewers as required. Templates could also be formatted to allow authors and/or peer reviewers to respond directly to individual concerns. Shared templates could assist with describing errors at post-publication review sites such as PubPeer and could facilitate the communication of published errors between PubPeer and journals. Structured templates could also encourage more standardized and transparent reporting in retraction notices and other post-publication reports (Vorland et al. 2020; Vuong 2020a, b), whereas universally accepted templates would allow information about

published errors to be aggregated, analysed and reported. This would improve the awareness and understanding of particular error types, which could encourage more journals to proactively investigate publications and submitted manuscripts for repeated errors such as wrongly identified nucleotide reagents. We propose a draft template (Table 4) that could align with and extend the “REAPPRAISED” checklist for publication integrity (Grey et al. 2020b). We welcome input from different stakeholders to ensure that this template is broadly useful and widely adopted.

While recognizing the challenges of communicating information about published errors, we note that a subset of retractions, expressions of concern and author corrections offered contrasting and at times inconsistent explanations for the incorrect control reagent. These different explanations for a common reagent error type may have also contributed to variations in journal responses. Indeed, contrasting explanations were unexpected, given that almost all (30/31) papers, and all 25 papers with published responses, stated that gene knockdown reagents were obtained from a common reagent supplier, Shanghai Hollybio. In 6/7 author corrections, the “non-targeting” control sequence was corrected to the same non-targeting control sequence, which is consistent with the support of a common supplier, and in the joint author correction of two papers, Holly Biotechnologies (or Shanghai Hollybio) accepted responsibility for the incorrect reagent description (Correction 2018). However, in contrast to assertions that the supplier error only affected the product description, the authors of other gene knockdown papers accepted that an incorrect non-targeting control had been used, or may have been used, and either requested or agreed to article retractions (Table 2). Taken together, these conflicting accounts indicate considerable uncertainty over the control reagents that were supplied and therefore the experiments that were conducted.

It was also surprising to note that same “non-targeting” control reagents were supplied by Shanghai Hollybio as well as other Shanghai-based companies. In total, 5 different companies were listed as supplying either Sequence A and/or Sequence D as “non-targeting” controls to human gene knockdown studies that analyse the functions of single human genes in human cancer cell lines. It is difficult to understand how multiple companies could supply identical incorrect reagents, or identical wrongly described reagents, when errors in products and/or their specifications might be reasonably expected to be manufacturer-specific (Knight 2001).

Regardless of whether errors affect material reagents or their descriptions, companies that supply incorrect reagents should make reasonable efforts to inform their clients of their mistakes (Knight 2001). Such responses seem particularly critical in the case of incorrect non-targeting control reagents that could be wrongly paired with targeting reagents for the analysis of literally thousands of human genes and hence provided to many individual clients. Published descriptions of specific incorrect non-targeting controls (Byrne and Labbé 2017; Byrne et al. 2019; Labbé et al. 2019) should also enable authors to contact reagent suppliers and/or take steps to correct their own publications, and we recognize that such efforts may be in progress. However, despite two years passing since the publication of author corrections where Holly Biotechnologies/ Shanghai Hollybio accepted responsibility for the incorrect description of a non-targeting control (Correction 2018), and despite numerous other papers that incorrectly describe Sequence A or Sequence D being clearly visible within the literature (Table 3), we could find no other author- or supplier-initiated published responses, beyond those analysed in the present study. It is increasingly difficult to understand why numerous papers that describe known incorrect control reagents should remain uncorrected within the literature.

Table 4 Draft communication template for published errors

| Field | Field title | Details/ explanation | Required per |
|-------|---|---|----------------------------------|
| 1 | Journal name | Prevents reporting to wrong journal | Report |
| 2 | How many individuals are reporting error(s) | Number automatically generates required number of fields 4–5 | |
| 3 | Number of papers to be reported/ journal | Number automatically generates required number of fields 6–8 | |
| 4 | Identity, ORCID, contact details of reporting party | Allow auto-population of contact details, nominate “corresponding contact” if more than one reporting party | |
| 5 | Description of reporting party | Drop down selection: author, reader, editor, peer reviewer | |
| 6 | PubMed ID | | Paper |
| 7 | DOI | | |
| 8 | Number of errors to report | Number automatically generates required number of fields 9–12 | |
| 9 | Error description, type of error | Drop down selections incorporating: fields from REAPPRAISED checklist ^a to add reagent description(s), allow free text response, allow file upload | Error or group of related errors |
| 10 | How error was identified and validated | Free text response, allow file upload | |
| 11 | Impact of error on reported results | Drop down selections, eg. affected results shown within article and/or supplementary data, error affects some/all results, allow free text response | |
| 12 | Impact of error on reported conclusions | Drop down selections, eg error affects all/ some conclusions, allow free text response | |
| 13 | Other papers with similar errors | Yes/ no | Report |
| 14 | Reported to PubPeer? | If yes, supply PubMed ID’s and/or DOI’s | |
| | | Yes/ no | |
| 15 | Other comments/ summary | If yes, link to PubPeer report | |
| | | Free text response | |

^aGrey et al. (2020b). Check for publication integrity before misconduct. *Nature*, 577(7789), 167–169

One previously highlighted journal retraction concerned a paper that described Sequence A as both a gene targeting and a non-targeting reagent and yet reported different results for these identical experiments (Byrne and Labbé 2017; Retraction 2016). The retraction notice stated that the “experiments were sourced out to a biotechnology company” (Retraction 2016). This retraction notice, combined with striking similarities between gene knockdown studies and their frequent shared nucleotide sequence reagent errors, led to the proposal that these cancer research papers may have been produced by undeclared assistance from external parties such as paper mills (Byrne and Labbé 2017; Byrne et al. 2019). Inconsistent explanations for the description of “non-targeting” control reagents, the implausible provision of identical “non-targeting” reagents by different external suppliers, and the apparently very limited efforts to correct affected papers by both reagent suppliers and study authors, could all be consistent with papers having been produced with undeclared external assistance, and with experiments not having been performed as described (Byrne and Labbé 2017).

Inconsistent explanations for the description of “non-targeting” control reagents also raise the concerning possibility that some information that is provided to journals in response to error notification could also reflect undeclared external assistance. This could mean that in addition to contributing papers to the literature, paper mills might also contribute information in response to notified errors or concerns, possibly to reduce the likelihood of publications being retracted. This possibility would considerably complicate journal efforts to appropriately respond to flagged errors and could result in the publication of what might be factually incorrect author “corrections”. Such incorrect corrections would allow study conclusions to remain unchallenged such that they could continue to mislead future cancer research efforts.

In summary, the present case study of journal responses to a set of human gene knock-down papers that share specific incorrect control reagents highlights the thin line that can exist between journal retractions, expressions of concern, author corrections, and taking no action at all. We propose that standardized error notifications might improve post-publication responses to published errors, enhancing the capacity of science to self-correct. Greater awareness and understanding of nucleotide sequence reagent errors should allow more journals to proactively identify and investigate papers that describe known incorrect reagents and take appropriate responses to papers that describe wrongly identified nucleotide sequence reagents. Indeed, the many cancer patients worldwide who look to biomedical research for improved treatments and quality of life deserve nothing less.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11192-021-03871-9>.

Acknowledgements We gratefully acknowledge Ms Jana Christopher and Dr Daniela Ruffell (FEBS Letters) for helpful discussions, and thank journal editors and editorial staff for their assistance and support.

Author contributions Conceptualization: Jennifer Byrne; Methodology: Jennifer Byrne; Formal analysis: Jennifer Byrne, Yasunori Park; Writing—original draft preparation: Jennifer Byrne, Yasunori Park; Writing—review and editing: Jennifer Byrne, Guillaume Cabanac, Amanda Capes-Davis, Bertrand Favier, Cyril Labbé, Yasunori Park, Rachael West; Funding acquisition: Jennifer Byrne, Amanda Capes Davis, Cyril Labbé; Supervision: Jennifer Byrne.

Funding JAB and CL gratefully acknowledge funding from the US Office of Research Integrity, grant ID ORIIR180038-01-00. JAB, CL and ACD gratefully acknowledge grant funding from the National Health and Medical Research Council of Australia, Ideas grant ID APP1184263.

Compliance with ethical standards

Conflicts of interest The authors have no relevant financial or non-financial interests to disclose.




Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). Reproducibility: a tragedy of errors. *Nature*, *530*(7588), 27–29.
- Bik, E.M., Casadevall, A., & Fang, F.C. (2016). The prevalence of inappropriate image duplication in biomedical research publications. *mBio*, *7*(3), e00809-16.
- Bik, E. M., Fang, F. C., Kullas, A. L., Davis, R. J., & Casadevall, A. (2018). Analysis and correction of inappropriate image duplication: the molecular and cellular biology experience. *Molecular and Cellular Biology*, *38*(20), e00309-e318.
- BioMed Research International. (2017a). Retracted: high expression of PTGR1 promotes NSCLC cell growth via positive regulation of cyclin-dependent protein kinase complex [retraction of: Biomed Res Int. 2016;2016:5230642]. *BioMed Research International*, 2017, 7640820.
- BioMed Research International. (2017b). Retracted: long noncoding RNA KIAA0125 potentiates cell migration and invasion in gallbladder cancer [retraction of: Biomed Res Int. 2015;2015:108458]. *BioMed Research International*, 2017, 3471417.
- Brown, A. W., Kaiser, K. A., & Allison, D. B. (2018). Issues with data and analyses: errors, underlying themes, and potential solutions. *Proceedings of the National Academy of Sciences USA*, *115*(11), 2563–2570.
- Byrne, J. A., Grima, N., Capes-Davis, A., & Labbé, C. (2019). The possibility of systematic research fraud targeting under-studied human genes: causes, consequences and potential solutions. *Biomarker Insights*, *14*, 1177271919829162.
- Byrne, J. A., & Labbé, C. (2017). Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines. *Scientometrics*, *110*, 1471–1493.
- Chiarella, P., Carbonari, D., & Iavicoli, S. (2015). Utility of checklist to describe experimental methods for investigating molecular biomarkers. *Biomarkers in Medicine*, *9*(10), 989–995.
- Correction. (2018). Correction to: cancer biother radiopharm 2014;29(8):303–309. DOI: <https://doi.org/10.1089/cbr.2014.1653> and Cancer Biother Radiopharm 2014;29(10):451–456. DOI: <https://doi.org/10.1089/cbr.2014.1698>. *Cancer Biotherapy and Radiopharmaceuticals*, *33*(8), 363.
- Fanelli, D., Ioannidis, J. P. A., & Goodman, S. (2018). Improving the integrity of published science: An expanded taxonomy of retractions and corrections. *European Journal of Clinical Investigation*. <https://doi.org/10.1111/eci.12898>.
- Georgescu, C., & Wren, J. D. (2018). Algorithmic identification of discrepancies between published ratios and their reported confidence intervals and P-values. *Bioinformatics*, *34*(10), 1758–1766.
- Grey, A., Avenell, A., Gamble, G., & Bolland, M. (2020a). Assessing and raising concerns about duplicate publication, authorship transgressions and data errors in a body of preclinical research. *Science and Engineering Ethics*, *26*(4), 2069–2096.
- Grey, A., Bolland, M. J., Avenell, A., Klein, A. A., & Gunsalus, C. K. (2020b). Check for publication integrity before misconduct. *Nature*, *577*(7789), 167–169.
- Guttinger, S., & Love, A. C. (2019). Characterizing scientific failure: putting the replication crisis in context. *EMBO Reports*, *20*(9), e48765.
- Habbal, W., Monem, F., & Gärtner, B. C. (2005). Errors in published sequences of human cytomegalovirus primers and probes: do we need more quality control? *Journal of Clinical Microbiology*, *43*(10), 5408–5409.

- Han, H. (2018). RNA interference to knock down gene expression. In J. DiStefano (Ed.), *Disease gene identification, methods in molecular biology* (Vol. 1706, pp. 293–302). New York: Humana Press.
- Katavetin, P., Nangaku, M., & Fujita, T. (2005). Wrong primer for rat angiotensinogen mRNA. *American Journal of Physiology Renal Physiology*, 288(5), F1078.
- Knight, J. (2001). When the chips are down. *Nature*, 410(6831), 860–861.
- Kocemba, K. A., Dudzik, P., Ostrowska, B., & Laidler, P. (2016). Incorrect analysis of MCAM gene promoter methylation in prostate cancer. *The Prostate*, 76(15), 1464–1465.
- Labbé, C., Cabanac, G., West, R. A., Gautier, T., Favier, B., & Byrne, J. A. (2020). Flagging errors in biomedical papers: to what extent does the leading publication format impede automatic error detection? *Scientometrics*, 124, 1139–1156.
- Labbé, C., Grima, N., Gautier, T., Favier, B., & Byrne, J. A. (2019). Semi-automated fact-checking of nucleotide sequence reagents in biomedical research publications: the seek and Blastn tool. *PLoS ONE*, 14(3), e0213266.
- Lambeth, L.S., & Smith, C.A. (2013). Short hairpin RNA-mediated gene silencing. In D. Taxman (Ed.), *siRNA design. methods in molecular biology (Methods and Protocols)*, vol 942 (pp. 205–232). Totowa: Humana Press.
- Malički, M., Utrobičić, A., & Marušić, A. (2019). Correcting duplicate publications: follow up study of MEDLINE tagged duplications. *Biochemia Medica (Zagreb)*, 29(1), 010201.
- Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48(4), 1205–1226.
- Retraction. (2016). Lentivirus-mediated TPD52L2 depletion inhibits the proliferation of liver cancer cells in vitro [Retraction]. *International Journal of Clinical and Experimental Medicine*, 9(6), 12416.
- Retraction. (2017). SUN1 silencing inhibits cell growth through G0/G1 phase arrest in lung adenocarcinoma [retraction of: *Onco Targets Ther.* 2017 Jun 02;10 :2825–2833]. *OncoTargets and Therapy*, 10, 5935.
- Saiz, L. C., Erviti, J., & Garjón, J. (2018). When authors lie, readers cry and editors sigh. *BMJ Evidence-Based Medicine*, 23(3), 92–95.
- Shannon, B. A., Cohen, R. J., & Garrett, K. L. (2008). Influence of 16S rDNA primer sequence mismatches on the spectrum of bacterial genera detected in prostate tissue by universal eubacterial PCR. *The Prostate*, 68(14), 1487–1491.
- Tamm, A. (2016). Incorrect primer sequences in the article on methylprednisolone treatment. *Acta Neurologica Scandinavica*, 134(1), 90.
- Vaught, M., Jordan, D. C., & Bastian, H. (2017). Concern noted: a descriptive study of editorial expressions of concern in PubMed and PubMed Central. *Research Integrity and Peer Review*, 2, 10.
- Vorland, C. J., Brown, A. W., Ejima, K., Mayo-Wilson, E., Valdez, D., & Allison, D. B. (2020). Toward fulfilling the aspirational goal of science as self-correcting: a call for editorial courage and diligence for error correction. *European Journal of Clinical Investigation*, 50(2), e13190.
- Vuong, Q. H. (2020a). Reform retractions to make them more transparent. *Nature*, 582(7811), 149.
- Vuong, Q. H. (2020b). The limitations of retraction notices and the heroic acts of authors who correct the scholarly record: An analysis of retractions of papers published from 1975 to 2019. *Learned Publishing*, 33(2), 119–130.
- Zhang, C., Chen, X., Li, Y., Himaya, S.W.A., Wu, J., Shi, X., et al. (2014a). si-RNA-mediated silencing of ADRBK1 gene attenuates breast cancer cell proliferation [published correction appears in *Cancer Biotherapy and Radiopharmacy*, 29(8), 303–309. *Cancer Biotherapy and Radiopharmacy*. 2018 Oct 1;33(8):363]. *Cancer Biotherapy and Radiopharmacy*, 29(8), 303–309.
- Zhang, C., Zhang, T., Song, E., Himaya, S. W., Chen, X., & Zheng, L. (2014b). Ribosomal protein S15A augments human osteosarcoma cell proliferation in vitro [published correction appears in *Cancer Biotherapy and Radiopharmacy*. 2018 Oct 1;33(8):363]. *Cancer Biotherapy and Radiopharmacy*, 29, 451–456.

Authors and Affiliations

Jennifer A. Byrne^{1,2}  · Yasunori Park²  · Rachael A. West^{2,3}  ·
Amanda Capes-Davis^{2,4}  · Bertrand Favier⁵  · Guillaume Cabanac⁶  ·
Cyril Labbé⁷ 

- ¹ New South Wales Health Statewide Biobank, New South Wales Health Pathology, Camperdown, NSW, Australia
- ² Faculty of Medicine and Health, The University of Sydney, New South Wales, Australia
- ³ Children's Cancer Research Unit, Kids Research, The Children's Hospital At Westmead, Westmead, NSW, Australia
- ⁴ CellBank Australia, Children's Medical Research Institute, Westmead, NSW, Australia
- ⁵ University Grenoble Alpes, Team GREPI, Etablissement Français du Sang, EA 7408, BP35, La Tronche, France
- ⁶ Computer Science Department, IRIT UMR 5505 CNRS, University of Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex 9, France
- ⁷ University Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France