



GABRIELA HASBUN FOR NATURE

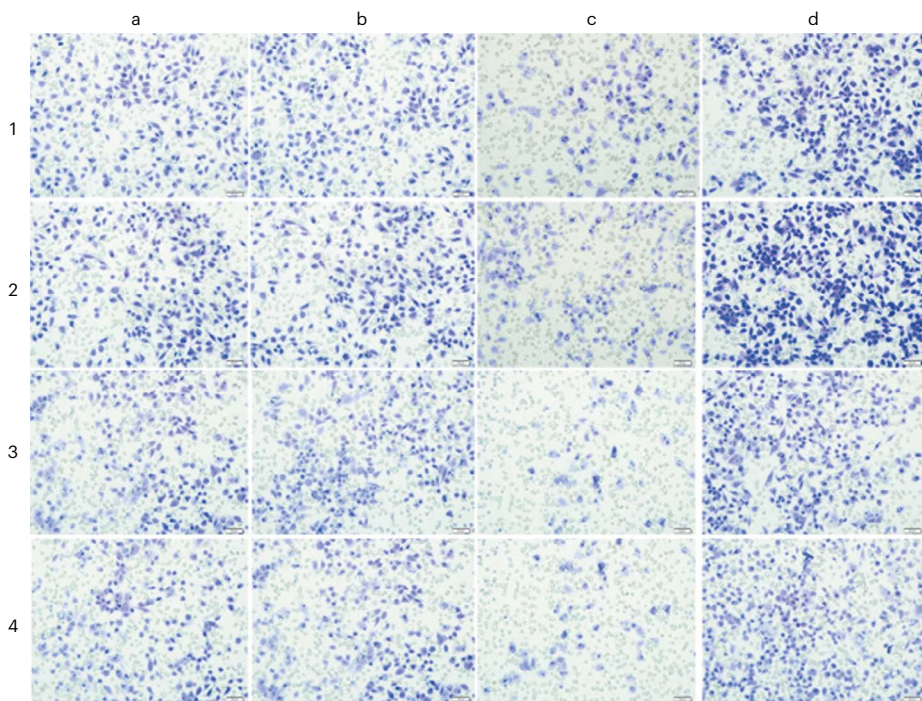
SEEING DOUBLE

Elisabeth Bik quit her job to spot errors in research papers – and has become the public face of image sleuthing. **By Helen Shen**



ARE YOU A SUPER-SPOTTER?

Elisabeth Bik identified five duplicated areas in these cell microscopy images from separate experiments in two figures in a paper. We have simplified the labels. Can you spot the duplications? (Answers overleaf).



SOURCE: S. GENG ET AL., PLOS ONE 9, E91566 (2014); RETRACTION 14, E0214018 (2019).

the Belgian message. Attached are images of western blots – the results of a common test to detect proteins in biological samples – from a published research paper. The writer wants to know: does Bik see anything fishy in this paper? Have these pictures been digitally altered?

Bik, a microbiologist from the Netherlands who moved to the United States almost two decades ago, is a widely lauded super-spotter of duplicated images in the scientific literature. On a typical day, she'll scan dozens of biomedical papers by eye, looking for instances in which images are reused and reported as results from different experiments, or where parts of images are cloned, flipped, shifted or rotated to create 'new' data (see 'Are you a super-spotter?'; answers overleaf at 'Did you spot them?').

Her skill and doggedness have earned her a worldwide following. "She has an uncommon ability to detect even the most complicated manipulation," says Enrico Bucci, co-founder of the research-integrity firm Resis in Samone, Italy. Not every issue means a paper is fraudulent or wrong. But some do, which causes deep concern for many researchers. "It's a terrible problem that we can't rely on some aspects of the scientific literature," says Ferric Fang, a microbiologist at the University of Washington, Seattle, who worked on a study with Bik in which she analysed more than 20,000 biomedical papers, finding problematic duplications in roughly 4% of them (E. M. Bik *et al.* *mBio* 7, e00809-16; 2016). "You have postdocs and students wasting months

or years chasing things which turn out to not be valid," he says.

Bik is not the world's only image sleuth, but she is unique in how publicly she presents her work. Many image checkers work behind the scenes, publishing their findings in research papers and writing privately to journals; a few are hired by journals or institutions. Some who flag up image problems work under pseudonyms, preferring not to be identified. But Bik posts her finds almost every day on Twitter and other online forums, in the process teaching others how to spot duplications and pressur-

"The things she calls out are usually real issues."

ing journals to investigate papers. In so doing, she's generated an "avalanche of reactions" and awareness about the problem, says Bucci. Bik estimates that her discoveries have led to at least 172 retractions and more than 300 errata and corrections – but all too often, she says, her warnings seem to be ignored.

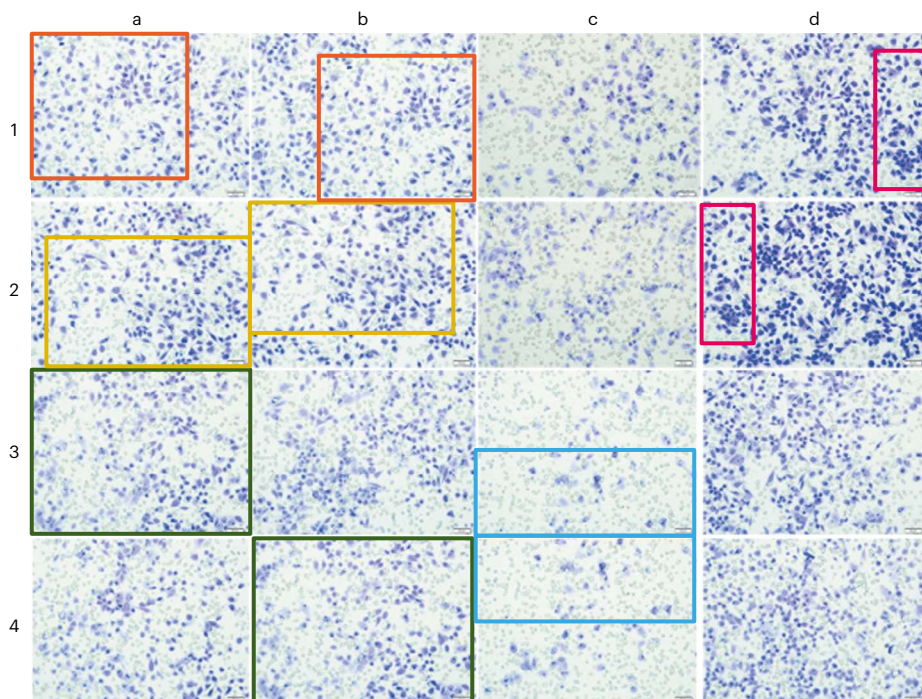
In April 2019, Bik announced that she had left her paid job at a biomedical start-up firm and would pursue image integrity work full-time, free of charge, for at least a year. A year on, she shows no signs of changing course – even though she has faced harassment, and at times been overwhelmed with requests. She's also shared her files with computer scientists trying to develop software to spot duplicated

February the fourteenth starts like most other days for Elisabeth Bik: checking her phone in bed, she scrolls through a slew of Twitter notifications and private messages from scientists seeking her detective services. Today's first request is from a researcher in Belgium: "Hi! I know you have a lot of people asking you to use your magic powers to analyse figures, blots and others but I just wanted to ask your opinion..."

After pouring a cup of coffee, Bik sits down at the long, wooden dining table that serves as her workstation at her home in California. She checks her e-mails on a giant 34-inch curved monitor, and takes a closer look at

DID YOU SPOT THEM?

Here are the duplicated sections Bik saw.



images across millions of papers, although the programs will probably always need human verification. “I’m enjoying it so much that I feel I just want to keep on doing this,” she says.

Hooked by a double smudge

Bik stumbled into image sleuthing around 2013, when, as a staff scientist at Stanford University in California, she read articles about scientific integrity and plagiarism. Out of curiosity, she googled quotes from her own published papers, and quickly found that other authors had lifted text without giving credit. “I was hooked. I was angry,” she says. “I immediately got fascinated about it, like how other people get fascinated by reading about crimes.” At one point, while examining a PhD thesis containing plagiarized text, something even more compelling caught her eye: a western-blot image with a distinctive smudge. The same image appeared in another chapter, supposedly for a different experiment. The chapters had also appeared as research articles, with the same errors, Bik saw. She e-mailed journal editors in January 2014; in June, she anonymously reported the papers online at PubPeer, a website where scientists can discuss published papers. These were Bik’s first reports of suspected manipulation in the literature. After an investigation by Case Western Reserve University in Ohio, the articles were retracted in 2015 and 2016.

Hunting for and cataloguing these images became a hobby. Then Bik contacted Fang and Arturo Casadevall, a microbiologist at Johns Hopkins University in Baltimore, Maryland. The trio decided that Bik’s rare talent could

lead to an in-depth inquiry of the frequency of problems in biomedical work. They sampled 20,621 papers, with Bik screening each – a task Fang says only she could do – before passing on her finds to Fang and Casadevall for corroboration. “It’s like a magic trick,” says Fang. “When it’s pointed out to you how it works, you can start to see it.” The team found 782 papers with what they termed “inappropriate” duplications, and Bik notified the relevant journals. The team reported the work in 2016 in *mBio*, at which Casadevall is editor-in-chief.

Bik spent so much of her spare time on duplicated images that last year she decided to leave her job as director of science at Astarte Medical in Foster City, California. “I realized I was getting more enthusiastic about image duplication work than my real job,” she says.

“It’s an impressive decision to make,” says Jennifer Byrne, a molecular biologist at the University of Sydney in Australia and herself a data-integrity sleuth who hunts for faulty genetic sequences in published papers. “It was very brave and, to be honest, pretty selfless.” Bik does not get paid for most of her work, but does some occasional paid consulting, and receives modest sums through a Patreon crowdfunding page. After decades of working and saving, she expects her current situation will be sustainable indefinitely.

The duplication database

Bik now operates out of a light-filled dining room, with floor-to-ceiling windows overlooking a garden filled with fruit trees and other plants, which she has catalogued in a spreadsheet. She also has a spreadsheet for

her collection of nearly 2,000 turtle figurines – gathered from travels and friends – which she keeps in a wall of glass cabinets. Most prized of all her spreadsheets, however, is a collection of more than 3,300 questionable papers, most of them flagged because of an issue with their images. (Bik sometimes raises other concerns with papers, such as around plagiarism or conflicts of interest.)

On a day without interruptions, Bik can peruse 100 papers or so, adding between 1 and 20 hits to her database (see ‘Advanced super-spotter test’; answers overleaf at ‘Did you spot them?’). A repeated smudge here or there, or a familiar smattering of data points: the visual indicators of duplication leap out at Bik from the screen. The collection is large enough to generate its own leads. It was looking at other papers by authors in her *mBio* data set, for instance, that led Bik last November to a case that generated her widest media coverage so far: a cluster of papers co-authored by Cao Xuetao, a prominent immunologist who has advocated for stronger research integrity in China, and who is the president of Nankai University in Tianjin. (Most of the articles listed Cao’s other affiliation, at the National Key Laboratory of Medical Immunology in Shanghai.) Bik and other pseudonymous commenters flagged apparent issues in more than 60 papers at PubPeer.

China’s ministry of education said it would investigate the articles, and Cao replied at PubPeer that he would re-examine the manuscripts, and that he was confident that the publications remained valid. Some authors replied swiftly on the site to point to honest errors. In one case, apparent duplicate images were in fact supposed to represent the same experiment but were not clearly labelled as such, an explanation that Bik accepts. In another, authors posted raw data and said the data seemed similar only after being processed for a paper. In still others, authors said there had been accidental mistakes, and by May this year, 13 of the flagged papers had received corrections, most stating that scientific conclusions weren’t affected. (Cao and China’s education ministry didn’t comment further for this article.)

Sometimes, Bik’s finds have pointed to suspected large-scale operations. This year, she and others have flagged a series of more than 400 papers that, they say, contain so many similarities that they could be the product of a ‘paper mill’ – a company that produces papers to order. Several image detectives worked to flag and collate the papers, including pseudonymous sleuths @mortenoxe, @TigerBB8 and @SmutClyde, who posted a list of papers in January, on a blog run by science journalist Leonid Schneider. “Finding these fabricated images should not rely solely on the work of unpaid volunteers,” Bik wrote in February on her own blog. Journals say they are now

SOURCE: S. GENGET AL., PLOS ONE 9, E91566 (2014); RETRACTION 14, E0214018 (2019).

investigating the papers, many of which are authored by doctors in Chinese hospitals, and some retractions are already being prepared.

Bik's data have revealed some insights into factors that correlate with image duplication. Her *mBio* paper reported that duplicated images had a slight tendency to occur more frequently in lower-impact journals. The paper also examined a subset of 348 articles flagged in *PLoS ONE*: taking into account the frequency of publication in the journal, it seemed that papers from China and India were more likely to contain problematic images. But Bik doesn't target one country's authors, she says. "I search for problematic papers, regardless of what country they are from," she wrote in November. In all, Bik has flagged up duplications in papers with lead authors from 49 countries.

Gamified sleuthing

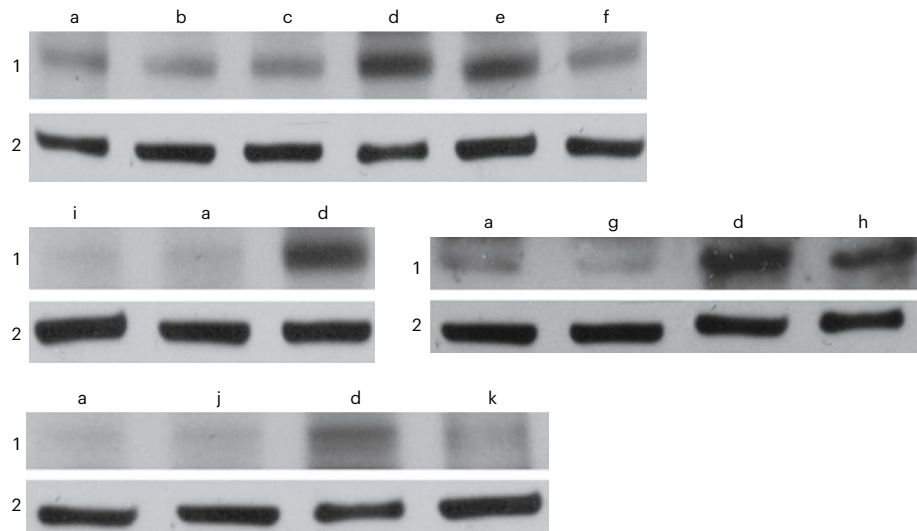
Nearly every day, Bik posts images with suspected problems to Twitter under the hashtag #ImageForensics, challenging her audience – which has almost tripled in the past year to more than 60,000 followers – to spot the matches before she posts her answers. The puzzles attract numerous guesses within minutes, and some eagle-eyed players spot issues that she misses. (She gives out emoji medals to top performers.) Bik says she hears from some followers who have picked up skills from her and spotted problematic images while peer-reviewing manuscripts. "I feel I'm changing people's way of looking at these images," she says. The work is sometimes overwhelming for Bik, who calls herself a "super introvert". Last November, she tweeted: "I am getting so many (anonymous) emails with people who want me to check certain authors or papers that I cannot possibly follow up. So many names ... And so much hidden pain among honest scientists about these dishonest coworkers."

Bik also posts detailed reports on what she sees to PubPeer, and occasionally comments there to support other tipsters. Many PubPeer users post their criticisms under pseudonyms – as does Bik in some cases, if she feels very worried about litigious authors. But she has posted more than 2,100 comments under her own name at the site since 2014. "What distinguishes Elisabeth is her willingness to identify herself, which is extremely admirable. It certainly helps with people taking the allegations seriously," says Mike Rossner, a former managing editor at the *Journal of Cell Biology* and president of Image Data Integrity, a consultancy firm in San Francisco, California.

Being unemployed and independent gives Bik the freedom to speak her mind, she says. "This one looks like nobody gave a fork about putting together a good science paper," she tweeted in March, with an accompanying figure panel that contained multiple duplicated

ADVANCED SUPER-SPOTTER TEST

Elisabeth Bik identified eight duplicate areas in these western blot images from separate experiments in four figures in a paper. We have simplified the labels. Can you spot the duplications? (Answers overleaf).



images. Last July, Bik commented on an image: "For those of you who did not get an NIH R01 grant around 2005, this is where that money was spent on instead."

But there is also risk, especially for someone who refers to herself as "blunt and snarky" on her Twitter biography. "At some point, I am afraid people will sue me," she says. She tries to keep her critiques to research papers, rather than accusing their authors. Bik has not faced a lawsuit, but has been harassed and has sometimes taken time off Twitter. One person e-mailed her former colleagues at Stanford arguing that she had abused her research grant funding by pursuing image integrity investi-

"We cannot, unfortunately, clone Elisabeth."

gations during work hours. (Bik says this was untrue.) Another posted personal information on PubPeer (now removed). "I've been called a bitch a couple of times," she says. "It comes with the work I do."

Because she posts under her real name, Bik says she errs on the side of caution, sometimes deciding not to flag cases online, especially those with blurry or low-resolution images. On her own science-integrity blog, many entries begin with some version of the phrase: 'This post is not an accusation of misconduct'. Suspicious images don't always point to corrupt actions, she says: researchers might have mistakenly uploaded a file twice when preparing figures, for instance. Then there are technical artefacts: membrane-thin slices cut sequentially from a piece of tissue can stick together along one edge and flip open butterfly-style, creating an apparent mirrored duplication.

Defects on an old microscope can create dark spots that seem the same on every image.

"She has a good track record," says Bernd Pulverer, chief editor of *The EMBO Journal*, who calls Bik a world leader in manual image screening. "The things she calls out are usually real issues."

Public or private

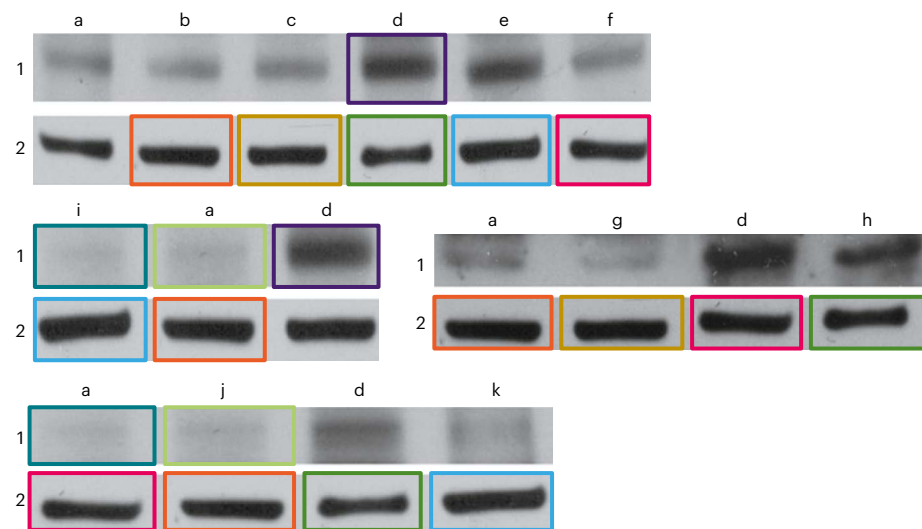
Although many praise Bik for her work, some say the concerns shouldn't be aired in public before they are flagged privately to journals or research institutions. "It's very problematic," says Lauran Qualkenbush, president of the US Association of Research Integrity Officers. She says that, in cases in which foul play is suspected, a public outing might hinder investigative procedures by universities. "If someone did conduct research misconduct intentionally, and then they're alerted to the concern, it's a great opportunity for them to destroy evidence," she says.

Bik – in common with other image sleuths – says she's tried informing journals privately, but the case often seems to go nowhere or take too long to resolve. (She also notes that researchers have opportunities to destroy evidence even if investigations occur in private.) Between 2014 and 2015, Bik reported all 782 questionable papers from her 2016 *mBio* study directly to journals through e-mail. Some journals were unprepared for the sheer volume of Bik's reports. She flagged 348 papers of concern to *PLoS ONE* in a raft of 30 e-mails, each with 10 or 20 attachments. "That obviously created a backlog because we were not equipped to deal with it," says editor-in-chief Joerg Heber. Eventually, in 2018, the journal formed a three-person team dedicated to investigating image integrity and other publication ethics cases full-time. "We published around 100 retractions last year.

SOURCE: Y. TAN ET AL., PLoS ONE 9, E102195 (2014); RETRACTION 14, E0220600 (2019).

DID YOU SPOT THEM?

In the original paper, the figures were far apart. Finding duplicates between papers is even harder.



Many of these were among cases that had been raised by her," says Heber. The team is still working through Bik's original tips, as well as other cases. Bik gives *PLoS ONE* credit for its efforts, and says she receives regular notifications of *PLoS ONE* retractions and corrections that have stemmed from her leads. But with many of the nearly 800 cases in the *mBio* data set still unresolved, Bik's patience is wearing thin. "I can tell you that 60–70% have not been addressed after five years, so now, yes, I'm going to take it more publicly," she says.

The due-diligence process to check concerns with papers often takes much longer than people expect, Pulverer says. He and Heber note that waiting for responses and raw data from authors, and sometimes research institutions, can be time-consuming.

Bik says she realizes that investigations take time. But she argues that journals could use expressions of concern more quickly and frequently to notify other researchers of potential problems, while possibly years-long investigations are pending. Heber says *PLoS ONE* uses expressions of concern when it has gathered enough information to be concerned, but might hold off if an investigation is running smoothly, in favour of reaching a resolution such as a correction or retraction. *Nature's* editor-in-chief, Magdalena Skipper, says that expressions of concern, which alert readers to "serious concerns" with a paper, are "a formal and permanent part of the scientific record; as such, we endeavour to use them judiciously, adding them to papers once we have evidence that it is appropriate to do so".

These days, Bik typically reports her discoveries directly on PubPeer. Some journals and publishers track activity on the site, so she can reach journal editors and the public. "It's more important to flag these papers and not worry about what happens behind the scenes with these institutes," she says.

Many – including Bik – argue that combating image manipulation and duplication requires system-wide changes in science publishing, such as greater pre-screening of accepted manuscripts. "My preference is not to have to clean up the published literature, but to do it beforehand," says Rossner. He helped to introduce universal image pre-screening of accepted manuscripts at the *Journal of Cell Biology* nearly 20 years ago. At the EMBO Press, says Pulverer, journals have pre-screened accepted papers for faulty images since 2013. But most journals still do not pre-screen or (as with *Nature*) spot-check only a subset of papers before publication. "Image screening is not common right now," says Chris Graf, director of research integrity at the publisher Wiley.

But the tide is slowly turning; Wiley publishes a few journals that screen images, and is "preparing to launch a screening service" with the *Journal of Cellular Biochemistry* and the *Journal of Cellular Physiology*, Graf says. The journal *Science* has editorial coordinators who check accepted manuscripts for signs of image manipulation, but they don't have capacity to check for some issues, such as whether figures have been flipped, rotated or duplicated, says executive editor Monica Bradford.

A job for AI?

Many researchers say automation is the key to improving image integrity at a large scale. "We cannot, unfortunately, clone Elisabeth," says Daniel Acuna, a computer scientist at Syracuse University in New York, whose group is one of a handful working on algorithms to detect problematic images. Although Bik excels at finding duplicated images in a single paper, computers could help to find more duplications between papers by comparing hundreds of thousands or millions of papers – an unfeasible task for humans, he says. In 2018,

Acuna's team published on the bioRxiv preprint server preliminary results of an analysis that extracted 2 million images from 760,000 papers (D. E. Acuna *et al.* Preprint at bioRxiv <http://doi.org/dtp2>; 2018). It proved too computationally intensive to compare every image with every other, but the team looked at image reuse within and across papers by the same authors. After manually examining a sample of more than 3,700 of the matching images that the software flagged, the researchers identified 40 cases that they all agreed were probably fraudulent; almost half of these involved the same image being used to represent different results in different papers.

Current technology is good at detecting outright duplications, and flipped or rotated copies, says Bucci. His company, Resis, uses proprietary software to scan scientific manuscripts for its clients, which include journals and research institutions. But complex problems are tougher, such as two images that share a small overlapping area, but are otherwise completely different. Advances in machine learning could be the key to detecting these and other subtle patterns automatically, he says.

But better software will need more data. Machine-learning algorithms require training with an abundance of images that are known to contain duplications. Bik has shared with Acuna images from hundreds of 'dirty' and 'clean' papers from her 2016 study. And at the Humboldt University of Berlin, researchers funded by the publisher Elsevier are developing a searchable database of images from retracted papers. For now, the collection has fewer than 500 entries, and most are in the life sciences and medicine and contributed by Elsevier, so the team wants more publishers to participate. The publisher says that some of its journals are piloting image-checking software, and its goal is to provide all its journals with automated systematic checking.

Until recently, Bik was unimpressed by the software available. Now, she says, "I have full confidence that in the next two years, computers will be usable as a mass way of screening manuscripts." But both Bik and Acuna say that people will always need to check the results of such programs, especially to weed out instances where images can and should look similar in certain parts.

For now, Bik has plenty of work to do. This morning's tip from Belgium looks like it might be a hit. Some of the western-blot bands – normally fuzzy and rounded like tiny black caterpillars – sport unusually sharp, pixellated edges, she says; these could be an innocent artefact introduced when a picture is compressed to a smaller size, or could suggest the application of photo-editing tools. "I'm going to ask him for the rest of the paper," says Bik.

Helen Shen is a science journalist based in Sunnyvale, California.

SOURCE: Y. TAN ET AL., *PLoS ONE* 9, E102195 (2014); RETRACTION 14, E0220600 (2019).