The Economist, Misconduct in Science: an array of errors, Sept. 10, 2011

ANIL POTTI, Joseph Nevins and their colleagues at Duke University in Durham, North Carolina, garnered widespread attention in 2006. They reported in the *New England Journal of Medicine* that they could predict the course of a patient's lung cancer using devices called expression arrays, which log the activity patterns of thousands of genes in a sample of tissue as a colorful picture. A few months later, they wrote in *Nature Medicine* that they had developed a similar technique which used gene expression in laboratory cultures of cancer cells, known as cell lines, to predict which chemotherapy would be most effective for an individual patient suffering from lung, breast or ovarian cancer.

At the time, this work looked like a tremendous advance for personalised medicine—the idea that understanding the molecular specifics of an individual's illness will lead to a tailored treatment. The team then started to organise a set of clinical trials of personalised treatments for lung and breast cancer. Unbeknown to most people in the field, however, within a few weeks of the publication of the *Nature Medicine* paper a group of biostatisticians at the MD Anderson Cancer Centre in Houston, led by Keith Baggerly and Kevin Coombes, had begun to find serious flaws in the work.

Dr Baggerly and Dr Coombes had been trying to reproduce Dr Potti's results at the request of clinical researchers at the Anderson centre who wished to use the new technique. When they first encountered problems, they followed normal procedures by asking Dr Potti, who had been in charge of the day-to-day research, and Dr Nevins, who was Dr Potti's supervisor, for the raw data on which the published analysis was based—and also for further details about the team's methods, so that they could try to replicate the original findings.

**A can of worms**

Dr Potti and Dr Nevins answered the queries and publicly corrected several errors, but Dr Baggerly and Dr Coombes still found the methods' predictions were little better than chance. Furthermore, the list of problems they uncovered continued to grow. For example, they saw that in one of their papers Dr Potti and his colleagues had mislabelled the cell lines they used to derive their chemotherapy prediction model, describing those that were sensitive as resistant, and *vice versa*. This meant that even if the predictive method the team at Duke were describing did work, which Dr Baggerly and Dr Coombes now seriously doubted, patients whose doctors relied on this paper would end up being given a drug they were less likely to benefit from instead of more likely.

Another alleged error the researchers at the Anderson centre discovered was a mismatch in a table that compared genes to gene-expression data. The list of genes was shifted with respect to the expression data, so that the one did not correspond with the other. On top of that, the numbers and names of cell lines used to generate the data were not consistent. In one instance, the researchers at Duke even claimed that their work made biological sense based on the presence of a gene, called *ERCC1*, that is not represented on the expression array used in the team's experiments.

Even with all these alleged errors, the controversy might have been relegated to an arcane debate in the scientific literature if the team at Duke had not chosen, within a few months of the papers' publication (and at the time questions were being raised about the data's quality) to launch three clinical trials based on their work. Dr Potti and his colleagues also planned to use their gene-expression data to guide therapeutic choices in a lung-cancer trial paid for by America's National Cancer Institute (NCI). That led Lisa McShane, a biostatistician at the NCI who was already concerned about Dr Potti's results, to try to replicate the work. She had no better luck than Dr Baggerly and Dr Coombes. The more questions she asked, the less concrete the Duke methods appeared.

In light of all this, the NCI expressed its concern about what was going on to Duke University's administrators. In October 2009, officials from the university arranged for an external review of the work of Dr Potti and Dr Nevins, and temporarily halted the three trials. The review committee, however, had access only to material supplied by the researchers themselves, and was not presented with either the NCI's exact concerns or the problems discovered by the team at the Anderson centre. The committee found no problems, and the three trials began enrolling patients again in February 2010.

Finally, in July 2010, matters unravelled when the *Cancer Letter* reported that Dr Potti had lied in numerous documents and grant applications. He falsely claimed to have been a Rhodes Scholar in Australia (a curious claim in any case, since Rhodes scholars only attend Oxford University). Dr Baggerly's observation at the time was, “I find it ironic that we have been yelling for three years about the science, which has the potential to be very damaging to patients, but that was not what has started things rolling.”

**A bigger can?**

By the end of 2010, Dr Potti had resigned from Duke, the university had stopped the three trials for good, scientists from elsewhere had claimed that Dr Potti had stolen their data for inclusion in his paper in the *New England Journal*, and officials at Duke had started the process of retracting three prominent papers, including the one in *Nature Medicine*. (The paper in the *New England Journal*, not one of these three, was also retracted, in March of this year.) At this point, the NCI and officials at Duke asked the Institute of Medicine, a board of experts that advises the American government, to investigate. Since then, a committee of the institute, appointed for the task, has been trying to find out what was happening at Duke that allowed the problems to continue undetected for so long, and to recommend minimum standards that must be met before this sort of work can be used to guide clinical trials in the future.

At the committee's first meeting, in December 2010, Dr McShane stunned observers by revealing her previously unpublished investigation of the Duke work. Subsequently, the committee's members interviewed Dr Baggerly about the problems he had encountered trying to sort the data. He noted that in addition to a lack of unfettered access to the computer code and consistent raw data on which the work was based, journals that had readily published Dr Potti's papers were reluctant to publish his letters critical of the work. *Nature Medicine* published one letter, with a rebuttal from the team at Duke, but rejected further comments when problems continued. Other journals that had carried subsequent high-profile papers from Dr Potti behaved in similar ways. (Dr Baggerly and Dr Coombes did not approach the *New England Journal* because, they say, they “never could sort that work enough to make critical comments to the journal”.) Eventually, the two researchers resorted to publishing their criticisms in a statistical journal, which would be unlikely to reach the same audience as a medical journal.

Two subsequent sessions of the committee have included Duke's point of view. At one of these, in March 2011, Dr Nevins admitted that some of the data in the papers had been “corrupted”. He continued, though, to claim ignorance of the problems identified by Dr Baggerly and Dr Coombes until the Rhodes scandal broke, and to support the overall methods used in the papers—though he could not explain why he had not detected the problems even when alerted to anomalies.

At its fourth, and most recent meeting, on August 22nd, the committee questioned eight scientists and administrators from Duke. Rob Califf, a vice-chancellor in charge of clinical research, asserted that what had happened was a case of the “Swiss-cheese effect” in which 15 different things had to go awry to let the problems slip through unheeded. Asked by *The Economist* to comment on what was happening, he said, “As we evaluated the issues, we had the chance to review our systems and we believe we have identified, and are implementing, an improved approach.”

The university's lapses and errors included being slow to deal with potential financial conflicts of interest declared by Dr Potti, Dr Nevins and other investigators, including involvement in Expression Analysis Inc and CancerGuide DX, two firms to which the university also had ties. Moreover, Dr Califf and other senior administrators acknowledged that once questions arose about the work, they gave too much weight to Dr Nevins and his judgment. That led them, for example, to withhold Dr Baggerly's criticisms from the external-review committee in 2009. They also noted that the internal committees responsible for protecting patients and overseeing clinical trials lacked the expertise to review the complex, statistics-heavy methods and data produced by experiments involving gene expression.

That is a theme the investigating committee has heard repeatedly. The process of peer review relies (as it always has done) on the goodwill of workers in the field, who have jobs of their own and frequently cannot spend the time needed to check other people's papers in a suitably thorough manner. (Dr McShane estimates she spent 300-400 hours reviewing the Duke work, while Drs Baggerly and Coombes estimate they have spent nearly 2,000 hours.) Moreover, the methods sections of papers are supposed to provide enough information for others to replicate an experiment, but often do not. Dodgy work will out eventually, as it is found not to fit in with other, more reliable discoveries. But that all takes time and money.

The Institute of Medicine expects to complete its report, and its recommendations, in the middle of next year. In the meantime, more retractions are coming, according to Dr Califf. The results of a misconduct investigation are expected in the next few months and legal suits from patients who believe they were recruited into clinical trials under false pretences will probably follow.

The whole thing, then, is a mess. Who will carry the can remains to be seen. But the episode does serve as a timely reminder of one thing that is sometimes forgotten. Scientists are human, too.

A link to Baggerly presentation (which includes a link to original data):

http://bioinformatics.mdanderson.org/Supplements/ReproRsch-All/Modified/StarterSet/baggerly\_nebraska12.pdf