

Impact case study (REF3b)

Institution: University of Warwick
Unit of Assessment: B10 Mathematical Sciences
Title of case study: Quality assessment for high-throughput genomic data in research and clinical practice
<p>1. Summary of the impact</p> <p>Dr Brettschneider and collaborators proposed a conceptual framework for high-dimensional gene expression data quality assessment (QA) and developed a QA statistical toolbox tailored to short oligonucleotide microarray technology. The work has deepened understanding of sources of variation and has helped in removing noise and bias in microarray data sets. This has accelerated the invention of clinical instruments for molecular cancer diagnosis/prognosis. The toolbox has been applied widely, leading to impact through:</p> <p>(A) process improvement in microarray facilities saving running costs, and standardisation of data quality targets ensuring reproducible research;</p> <p>(B) individualisation of treatment decisions supported by enhanced data quality, thereby reducing healthcare costs through avoidance of unnecessary surgery and improved patient welfare.</p>
<p>2. Underpinning research</p> <p>Microarray based research. Genome-wide expression profiles have become central to understanding the genetic underpinning of cellular processes (<i>e.g.</i>, cell division, circadian rhythm, reaction to chemicals) and complex genetic diseases (<i>e.g.</i>, cancers, neuropsychiatric disorders). This approach was facilitated by the invention of microarrays two decades ago, which radically changed gene expression measurement. Instead of assessing genes one-by-one, the new technologies enabled the screening of tens of thousands of them at once.</p> <p>Reproducibility of research in genomics. After initial euphoria, many microarray studies turned out to be inconclusive or irreproducible, affecting industrial research in particular. In 2011, Bayer, one of the world's largest chemical companies, halted nearly two-thirds of its target-validation projects because in-house experimental findings failed to match published literature claims; see blogs.nature.com/news/2011/09/reliability_of_new_drug_target.html. In the last decade, the need for rigorous experimental design, for appropriate statistical interpretation of results and for collaboration with statisticians has become widely recognised in the genomics community. <i>Nature</i> and other related research journals have published a series of articles about reproducibility and have taken editorial steps to ensure transparency and robustness in articles published there; see nature.com/nature/focus/reproducibility.</p> <p>QA for high-throughput gene expression data. In a seminal paper [1] including five commentaries from researchers inside and outside academia, Brettschneider <i>et al</i> laid the groundwork for a new branch of research in statistical genomics. On the theoretical side, they provided a general conceptual framework for quality assessment (QA) for data obtained by high-throughput molecular assays such as gene expression microarrays. On the practical side, they proposed a statistical QA toolbox for such gene expression measurement technologies. The methods include numerical chip quality measures such as <i>RLE and NUSE distributions</i>, as well as spatial representations of by-products of pre-processing algorithms, also called <i>quality landscapes</i>, which enable scientists to connect technical artefacts to location on the slide. All measures were studied extensively on a variety of datasets, including spike-in calibration data, small lab and multisite studies. Some of the tools are specific to short oligonucleotide microarrays (including the industrial standard made by Affymetrix); others can be used for different microarray platforms or even for the most recent gene expression measurement technologies such as RNAseq.</p> <p>The research team. Brettschneider joined the University of Warwick in 2007 and collaborated on [1] with Collin (Genomic Health, previously Affymetrix), Bolstad (Affymetrix) and Speed (UCB & WEHI). The work reported in [1] had begun in 2006 while Brettschneider was working at Queen's University in Ontario, and was completed in 2008 (with the aid of a visit to Warwick, from California, by her co-author Prof Speed).</p>

3. References to the research

Key publication: The bulk of the work was published in *Technometrics*, the leading journal for statistical methodology in sciences and engineering. The editors invited five commentaries and a rejoinder, indicating the central nature of the work for the field of statistics in genomics and beyond. Brettschneider was invited to present the paper at JSM (*Joint Statistical Meeting* of the American Statistical Association), the world's largest gathering of statisticians, attracting over 5000 speakers and participants from academia, research institutes, government and the commercial sector, both from within USA/Canada and overseas.

[1] Primary article: Brettschneider J, Collin F, Bolstad BM, and Speed TP, Quality assessment for short oligonucleotide arrays, *Technometrics* **50** (2008) 241-264.

(With five commentaries: Jones W, Bao L & Hoeschele I, He W & Bull SB, Kendall J & Lakshmi B, Goldstein D, ditto pp.265-278, and Rejoinder, Brettschneider J et al, ditto pp.279-283).

All parts available at pubs.amstat.org/toc/tech/50/3 Cited 53 times to 16.10.2013.

[1] was included, as an *Internet Publication* (*arXiv* preprint), in Warwick's RAE 2008 submission.

Software: The statistical QA tools from [1] have been incorporated into software packages (see next Section and reference [2] for details).

4. Details of the impact

The main route for delivering impact for the statistical QA tools of [1] is via a number of software packages [2]. These include affyPLM and affyQualityMetrics, available at bioconductor.org, a repository for free open source bioinformatics software, started in 2001 by an international team led by members of the Fred Hutchinson Cancer Research Center, and now the default software resource for the bioinformatics community. The reach of this impact is considerable. For example, during the REF period, the two packages have been downloaded from Bioconductor over 380,000 times [2 i, ii]; Chipster's server in Finland (one of ten such servers worldwide) has 788 users [2 iii]; and the Genevestigator web tool which uses the Bioconductor package affyQCReport (incorporating affyPLM) for the QC of datasets is used for [text removed for publication] [2 vi]. With two of the co-authors working for biotech companies it was easy to establish further connections to the commercial world and spread the word about the methodology. Invitations to conferences such as JSM helped in sharing the methods, both with other statisticians and with users. More details about the routes to impact are given in the explicit examples below.

(A) Process improvement in microarray facilities and standardisation of data quality targets

The availability of QA tools [1] has enabled large microarray facilities as well as smaller laboratories to routinely monitor and improve their processes. This enables them to use high quality data for downstream analysis, ensuring validity of scientific conclusions. Two supporting letters [3, 4] confirm this:

[3] CEO of Spheromics (formerly at Novartis and the Australian Genome Research Facility):

"Quality assessment and control for microarray data is an area where existing quality control methodology was not sufficiently developed [...]. The methodology [from [1]] has been helpful to me for both detecting outlier arrays and revealing systematic errors in the process. [...] I have been using the tools proposed and the guidance about their interpretation, in particularly the NUSE distributions. They would highlight, at a glance, differences of chip quality caused by batch effects."

[4] Senior Bioinformatics Officer at ICR (Institute of Cancer Research):

"We use quality assessment based on statistics methods developed by [1] implemented in R-packages. Quality assessment has been helpful to us in determining outlier arrays and detecting artefacts in the data. This makes sure our dataset is not distorted by poor quality microarrays."

The QA methods from [1] have contributed to changing attitudes in the genomics community, both in academia and beyond, towards giving higher priority to routine checking of data quality and to understanding of technically caused variation before drawing scientific conclusions. For example:

- The *quality landscapes* feature provide a visual presentation of quality defects, thereby allowing scientists to identify, in the measurement process, specific causes of poor quality. The popularity of [5] (e.g., 10715 visits in the 12 months preceding February 2013) confirms the usefulness of this

Impact case study (REF3b)

approach.

- Making use of the QA tools for microarray data has become a routine step in scientific studies to ensure the reproducibility of research findings; see e.g. [6] for one of many examples. Many authors cite the software or names of the QA tools (RLE, NUSE) rather than the publication [1].

The methods [1] have played important roles in setting professional standards for microarray data quality, as illustrated by an international initiative led by the US Food and Drug Administration (FDA). The Microarray Quality Control (MAQC) project aims at establishing standards to ensure successful and reliable use in clinical practice and regulatory decision-making. Phase II aimed to assess and establish “best practices” for development and validation of predictive models for personalised medicine. CEO of Spheromics, who was involved in the project, states in his letter [3] *“In Phase II of that project, I have made frequent use of the assessment methods from your paper.”*

(B) Individualisation of treatment decisions

In the last decade, many biotech companies have been pushing for the development of diagnostic and prognostic tools to support the optimisation of treatment choices. For example, individualised recurrence estimates help decide whether or not a patient would benefit from adjuvant chemotherapy. Clearly the need for high quality standards becomes increasingly important the closer the use of a tool gets to a clinical setting. The QA methods developed in [1] are being used in the research units of companies developing tools for personalised medicine.

We demonstrate this in detail for Veracyte, a molecular diagnostics company and a pioneer in the emerging field of molecular cytology. Veracyte has developed a gene expression based test called Afirma that is expected to reduce massively the number of surgeries with their attendant morbidity (life-long follow-up treatments) in initially suspected thyroid cancer due to the occurrence of fairly common thyroid nodules. A recent economic impact study in the Journal of Clinical Endocrinology and Metabolism concluded that routine use of Afirma in the USA would result in 74% fewer surgeries in patients with benign tumours, that is, tens of thousands of avoidable surgeries each year corresponding to about \$122 million medical savings [7]. The traditional diagnosis of thyroid cancer produces up to 30% inconclusive cases which typically result in surgery, of which 70%-80% of patients turn out to have benign tumours. Afirma succeeds in avoiding the need for surgery in about half of such cases, resulting in expected health care cost savings of \$3000 per patient as well as improving patient health outcomes. Afirma has been developed and clinically validated as published in 2012 in high-profile journals such as the New England Journal of Medicine (*NEJM*) [8].

A crucial step for the commercial success of Afirma is to obtain medical insurance cover, e.g., it was approved by Medicare in 2012. For this, FDA software validation is key, a step that takes six months and about \$300,000 and after which the algorithm is locked. To ensure their success in winning over the clinicians, Veracyte needed the negative predictive value of their tool to be above 94%, which required the highest possible quality standards. Veracyte’s Chief Scientific Officer states in her supporting letter [9]:

“A recent study found that the Afirma Gene Expression Classifier demonstrated strong accuracy, reliability and reproducibility under a range of conditions and variables. This includes maintenance of sample stability, analytic sensitivity and analytical specificity [...] A key step to these achievements is data quality assessment and control. We have been using quality assessment methods such as summaries of RLE distributions from your publication to shed light on the sources of variation in our custom-made gene expression microarrays. [They] have been serving us for detecting outliers as well as for revealing and removing artefacts and batch effects arising from inconsistencies in operator, protocol or sample condition.”

Several media stories from 2012 onwards highlight how particular patients, after having taken the Afirma test, have avoided surgery with long term health savings and improved patient welfare [10].

5. Sources to corroborate the impact

[2] QA tools from [1] have been incorporated into numerous software packages including:

(i) Bioconductor R-package *affyPLM* for preprocessing Affymetrix type data. Part III of the package vignette is devoted the QA tools; downloaded 344,040 times in REF period:

Impact case study (REF3b)

bioconductor.org/packages/release/bioc/html/affyPLM.html

(ii) Bioconductor R-package *arrayQualityMetrics* described in Kauffmann *et al*, *arrayQualityMetrics* – a bioconductor package for quality assessment of microarray data, *Bioinformatics* 25 (3): 415-416; downloaded 37 times in 2009, 180 times in REF period:

bioconductor.org/packages/release/bioc/html/arrayQualityMetrics.html

Correspondence received providing download statistics.

(iii) *Chipster* described in M Aleks Kallio *et al*, *Chipster: user-friendly software for microarray and other high-throughput data*, *BMC Genomics* 12:507, 2011: chipster.csc.fi/index.shtml

Correspondence received corroborates users.

(iv) *RobiNA* described in Lohse M *et al*, *Robin: An intuitive wizard application for R-based expression microarray quality assessment and analysis*. *Plant Physiology*, 153:642-51, 2010.

Incorporates all QA tools from [1], see Section 5.1 in mapman.gabipd.org/web/guest/robin

(v) Maastricht Bioinformatics has built an online Affymetrix array QA webtool running from arrayanalysis.org which includes NUSE and RLE from [1].

(vi) Nebion Gene expression search engine explains the use of QA from [1] in Chapter 8 at genevestigator.com/userdocs/manual/gc.html. [text removed for publication]

[3] Letter received: Founder and CEO of Spheromics spheromics.com, a company specialising in consultancy for biomarker development and gene expression.

[4] Letter received: Senior Bioinformatician at ICR: Royal Cancer Hospital, London.

[5] Catalogue of technical artefacts: The website “Chip Gallery” plmimagegallery.bmbolstad.com set up by Bolstad provides QA case studies with extensive collections of quality landscapes. The emphasis is on spatial quality effects which lab scientists can compare with their own experimental results to assign causes of poor quality. 10715 visits in the 12 months preceding February 2013.

[6] Influencing methodology in the user community: Zhao H and Ma H, *FacPad: Bayesian sparse factor modelling for the inference of pathway responsive to drug treatment*. *Bioinformatics* (2012) 28 (20): 2662-2670 bioinformatics.oxfordjournals.org/content/28/20/2662.short. The methods section reports all analysis steps starting with the raw data modelling transparency: “...basic quality check of the .CEL data was performed using the NUSE and RLE metrics (Brettschneider *et al.*, 2008) [i.e. [1]], which are provided in the R package ‘affyPLM’. [...] boxplots of NUSE and RLE were drawn and arrays with bad quality were discarded.”

[7] Li H *et al*, *Cost-Effectiveness of a Novel Molecular Test of Cytologically Indeterminate Thyroid Nodules*, *JCEM* 96(11): E1719 icem.endojournals.org/content/96/11/E1719.full.pdf+html

[8] Alexander EK *et al*, *Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology*, *NEJM*, Aug 23, 2012 nejm.org/doi/full/10.1056/NEJMoa1203208#t=abstract

[9] Letter received: Chief Scientific Officer, Senior Vice President of Research and Development at Veracyte veracyte.com, Bay Area Biotech pioneering in molecular cancer prognosis, among top 50 Venture-capital companies by Wall Street Journal’s *Next Big Thing 2011*.

[10] Examples of news stories in US media highlighting that patients have benefitted directly from the Afirma test which relies on [1] as corroborated by [9]:

Procedure at local doctor's office helps local woman avoid thyroid surgery (March 2012)

heraldnews.com/news/x570358499/Procedure-at-local-doctors-office-helps-Fall-River-woman-avoid-thyroid-surgery

New Thyroid Cancer Test (December 2012)

abclocal.go.com/kfsn/story?section=news/health/health_watch&id=8918522

Thyroid cancer test helps patients avoiding unnecessary surgery (May 2013) where a clinic notes several patients have benefitted from the test:

desmoinesregister.com/article/20130520/LIFE/305200013