available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Selective discussion and transparency in microarray research findings for cancer outcomes

John P.A. Ioannidis[a,b,*], *Nikolaos P. Polyzos*[a], *Thomas A. Trikalinos*[a,b]

[a]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece*

[b]*Department of Medicine, Tufts University School of Medicine, Boston, MA, USA*

ARTICLE INFO

Article history:
Received 20 March 2007
Received in revised form 22 May 2007
Accepted 23 May 2007
Available online 12 July 2007

Keywords:
Microarrays
Survival
Cancer
Prediction
Reporting
Databases
Discussion

ABSTRACT

We examined the interpretation of research findings and public availability of transparent information on data and processing for 46 articles of microarray studies that had addressed major cancer outcomes. Unsupervised and supervised methods selected molecular signatures with a median of 675 and 50 genes, respectively, but only a median of eight genes or groups thereof were further discussed. Across 479 genes or groups thereof discussed in all 46 studies, 65% reflected specific comments (reflecting external relevant data from other studies or other lines of reasoning relevant to the gene of interest), and 59% of the comments were referenced. Among specific comments, supportive ones outnumbered comments against the research findings by nine to one (270 versus 29). Discussion was similarly selective in early studies and in studies published in 2006. Even in 2006 only 10 of 15 studies had publicly deposited data. Only three studies had scanned images, raw and processed data available. Processing details varied. Public transparency and unbiased interpretation of findings can be improved in microarray research.

1. Introduction

Gene expression profiling with microarrays is a prime paradigm of high-dimensional biology where the resulting signatures typically include many gene tags.[1] Molecular signatures have been proposed that may provide important information for major clinical outcomes, in particular for several cancers.[2] However, the reliability of molecular signatures depends on careful attention to experimental detail.[3] The complexity of the massive testing process requires transparency at all stages of microarrays experiments.[4,5] Data and their processing should be reproducible and publicly available.[6–8] Suboptimal design and selective reporting may diminish the credibility of the findings.[9–11] An additional concern is that, when a molecular signature is finally derived, making inferences about why some genes were included in the signature while others were not remains a challenge.

Here we aimed to examine whether there is evidence for selection bias in the interpretation of microarray results and whether this is accompanied by deficiencies in the public data availability and transparent processing. We considered two sets of studies[12–57] that generated gene expression profiling for prediction of major clinical outcomes (survival, disease-free survival, or response to treatment) in malignancies: those published in the early years (1999–February 2003) when microarrays were first applied as tools to generate prognostic and predictive signatures for cancer ($n = 31$) and those that were very recent (published in the first half of 2006, $n = 15$). These datasets also allowed a comparative examination on whether any improvements had occurred over time.

* *Corresponding author.*
E-mail address: jioannid@cc.uoi.gr (J.P.A. Ioannidis).
0959-8049/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.ejca.2007.05.019

2. Materials and methods

2.1. Database of analysed studies

We used two databases of studies that used microarrays to classify malignancies into subtypes. First, we employed a previously assembled database of such studies published until February 2003,[2] focusing on the 31 studies that attempted to predict major clinical outcomes (survival, disease-free survival, or response to treatment) and that evaluated at least ten patients with cancer using cDNA or oligonucleotide analyses on at least 500 genes. Studies were included regardless of whether they had also tried to generate molecular signatures for additional disease correlates and phenotypes (e.g. stage or histology) or not, and regardless of how many gene probes were eventually selected in the molecular signatures. Studies assessing differential gene expression between different malignancies or between patients with malignancy and cancer-free controls were excluded. We also excluded publications that re-analysed already published datasets without adding new cases. Second, we used the exact same PubMed search and inclusion-exclusion criteria to identify the 15 eligible studies that were published in the first 6 months of 2006. The two databases also allowed us to compare whether any features of the discussion and interpretation as well as public data availability and transparent processing in these studies had changed between the early and more recent studies. For public data availability, MIAME-compliance and transparent processing, we did not consider studies published before June 2002, since the MIAME guidelines were first published in December 2001[4] and further promoted in the summer of 2002,[58,59] so it would be unfair to require much detail and data transparency of studies published before even the early promotion of MIAME.

2.2. Evaluations and data

From each study we recorded the first author's name, journal and year of publication; the type of cancer; the major outcomes for which molecular signatures were developed; whether the selection of the subtypes used approaches dependent or independent on the outcome (supervised versus unsupervised clustering); the number of genes in the molecular signature(s) selected; and whether independent validation had been performed or not.

We further scrutinised each article for discussion comments to individual genes/probes or sets thereof. For operational purposes each mention of a gene in the text was considered a 'comment'. For articles having a distinct Discussion section, we scrutinised that section only. For articles that did not separate the Discussion section from the Results, we scrutinised the combined Results and Discussion. For articles with no structure, we scrutinised the whole text for any discussion comments. We excluded comments that were included in technical appendices.

For each gene or group of genes mentioned in the discussion-relevant sections, we recorded whether the comment was 'specific' or 'non-specific'. We defined as 'specific' all comments that reflected external relevant data from other studies or other lines of reasoning relevant to the gene of interest (i.e. how the gene fits into a molecular pathway that explains or refutes the observed results). All other comments were classified as 'non-specific'. Non-specific comments include simple mentions of the gene without any further comment, re-iteration of the study's results, or mere descriptions of the gene's usual function without any clear reasoning or hypothesis. We further recorded whether the comment was referenced or not. Comments referring to specific data or appealing to specific streams of reasoning were further classified as being clearly supportive of the gene finding ('for'); clearly against the finding ('against'); or neutral/unclear. The categorisation was performed as follows. When specific data from other studies were mentioned, we recorded whether they presented a biological effect in the same direction as the biological effect found in the microarrays paper ('for'); a biological effect in the opposite direction than in the microarrays paper or no biological effect ('against'); or the biological effect was not stated or was unclear. When a stream of reasoning was employed rather than specific data, we decided whether the arguments were for or against the finding of the microarrays paper or whether they were neutral/unclear. The 'against' category included also cases where the discussion mentioned that a specific gene had been found to be important in some other study, but was not part of the molecular signature identified in the current study.

Whenever specific external data or a specific line of reasoning was mentioned in the discussion, we also appraised their apparent relevance. Apparent relevance was graded with respect to the commented gene and with respect to the disease under study. Thus, we classified comments pertaining to the exact same gene versus all other comments (i.e. those that pertained to a different gene, a broader set of genes that included the gene of interest, or comments pertaining to linkage to a chromosomal region). We also categorised comments into two groups with respect to the disease to which they pertained: same disease; or other phenotypes and diseases (i.e. including other types of cancer; other disease thought to involve the same pathway; *in vitro* cell lines or animal models; or any other diseases).

The appraisal and categorisation of comments was performed for all studies and also separately limited to studies that had included supervised approaches, because one may argue that unsupervised clustering methods do not primarily aim to generate molecular signatures per se with specific genes being implicated.

For each study, we also scrutinised the published article and online supplements to identify if there was any mention of public availability of datasets and protocols/methods. We identified each alluded publicly-accessible dataset and accessed it for further scrutiny. We recorded for each publicly accessible dataset whether data were available for the scanned images; for raw data; and for processed data (normalised and parameterised or standardised data used eventually for analysis). We also recorded whether these data were compliant with MIAME specifications.[4] We scrutinised whether the array platform was available on public access with details in the description of each array element, as specified by MIAME.[4] Furthermore, we examined the information that was provided either in the article and supplements or in the public dataset records for details on the procedure of image scanning and the normalisation and standardisation processing of the data. We examined whether information on

processing would suffice to generate the processed data from the raw data, when only the raw data were provided in the public databases. Finally, for studies where the processed data were available, we perused the presented information on the statistical analysis employed, noting whether software were described; and whether statistical and bioinformatics methods were described; and if so, in what detail.

2.3. Data extraction process and analysis

Two investigators undertook independent data extraction using pre-constructed forms (TAT and NPP for discussion comments; and JPAI and NPP for all other aspects). All papers were examined in duplicate. A pilot form was used to data-extract five randomly chosen papers; after a consensus meeting of all investigators, amendments were made to the form, definitions were clarified in detail, and the remaining papers were scrutinised. The independent data extractions were compared. Remaining discrepancies were settled with a third investigator. Consensus was reached on all items.

Operational examples (developed before data extraction to guide the process) and real examples on discussion comments are shown in Appendix Table A.1. As compared with the final consensus, the two independent data extractors had agreed upfront (before consensus) on 94%, 88% and 95% of genes, 'specific' data/reasoning, and use of reference items, respectively. All of the comments not in line with ('against') the observed effects were identified by both data extractors, whereas data extractors agreed 92% on supporting ('for') comments.

We present descriptive information on the 46 studies, as defined above. We also compared the group of early studies versus the group of recent studies using Fisher's exact test and Mann–Whitney U test for binary and continuous variables respectively. P values are two-tailed.

3. Results

3.1. Characteristics of analysed studies

Table 1 shows characteristics of the 46 analysed microarray studies. The molecular signatures in 19 of the 46 analysed studies were also subjected to some form of independent validation, more commonly in recent rather than earlier studies (11/15 versus 8/31, exact $P = 0.004$). Unsupervised clustering typically selected very large numbers of genes (between 100 and 11322 genes, with five exceptions where only 10–90 genes were selected, median 675). Molecular signatures based on supervised approaches selected fewer genes than unsupervised methods; however, the molecular signatures for the major clinical outcome still contained a median of 50 genes (range 5 to 512). The total number of selected genes could actually be higher when additional disease correlates and phenotypes were studied (Table 1), but then it was not possible to calculate the total number of non-overlapping gene probes.

Seven of the 46 studies did not discuss anything about any gene among those included in the identified molecular signatures. The median number of genes or groups thereof with comments of any sort (including simple mention of a gene and/or its function in the text of the Discussion) was only eight (Table 1).

3.2. Comments on genes and groups thereof

Overall, 479 genes or groups thereof were discussed across all the 46 articles. Of those, the discussion comments reflected specific data or a specific stream of reasoning for 314 (65%, 95% confidence interval 61–70%) genes or groups thereof in 37 papers; two of the 39 papers that commented on some genes had only non-specific comments. References were used in a similar proportion ($n = 285$, 59%, 95% confidence interval 55–64%). There was high concordance between referencing and specificity of the comment (kappa concordance coefficient 0.60, $p < 0.001$) (Table 2).

Thirty six of the 37 studies that had at least one 'specific' comment made comments that were in line with ('for') the observed effects in the microarrays study (Table 3). Only nine studies made comments that were 'against' the observed effects and 8 of them also made at least as many or more comments 'for' the observed effects. Only one recent study made solely comments 'against' the observed direction. Across the database, supportive comments outnumbered comments that were not in line with the findings by more than 9 to 1 (270 versus 29). Of the 29 comments 'against' the findings, 26 pertained to the situation where there was no change in the expression of the commented gene contrary to expectations based on previous studies or biological plausibility; only three comments pertained to the situation where the current molecular signature included a gene that had not been found to be important in other lines of evidence. Neutral/unclear comments (total $n = 15$ comments) were made in seven studies.

Of the 314 comments that referred to specific external data or specific stream of reasoning, 262 (83%) used a rationale appealing to the same gene (Table 3). In the remaining cases, the comment pertained to a broader category of genes ($n = 33$), the wider chromosomal region ($n = 18$), or even a different gene ($n = 1$). Only in 147 (47%) comments was the disease phenotype clearly the same (Table 3). In the remaining cases, the comment pertained to a broader pathway ($n = 104$), a cell line or animal model ($n = 18$), a different cancer or cancer generally without specifying histological type ($n = 42$), or even another non-malignant disease ($n = 3$).

A similar picture was seen when analysis of comments was limited to the studies that had performed supervised clustering. Among 37 studies, 393 genes or groups thereof were discussed. Of those, 251 (64%) reflected specific comments, and 219 (56%) of the comments were referenced. Among specific comments, supportive ones outnumbered comments against the research findings by over eight to one (219 versus 26; and six neutral comments) and comments on the same gene and same disease phenotype represented 83% (209/251) and 44% (111/251), respectively.

3.3. Early versus recent studies

We further examined whether the situation is different in the more recent studies compared with articles describing earlier efforts in the field (Table 4). All 15 recent studies commented on at least one gene or group of genes, while seven of the 31 early articles had not ($P = 0.08$). The percentage of specific or referenced comments among all comments, the percentage of specific comments that were 'for' the research finding,

Table 1 – Evaluated microarray gene expression studies for major cancer outcomes

Author (year)	Neoplasia	Major clinical outcome	Selected genes for major outcome	Other correlates examined	Comments on genes or groups thereof
Early studies					
Ahr (2002)	Breast	Metastatic disease	41 (U)	Yes	None
Alizadeh (2000)	DLBCL	Death	100 (U)[a]	No	None
Beer (2002)	NSCLC	Death	4966 (U), 50 (S)[a]	Yes	9
Belbin (2002)	Head and neck cancer	Death	375 (U)	Yes	21
Bertucci (2002)	Breast	Death	1045 (U)	Yes	25
Bhattacharjee (2001)	Lung adenocarcinoma	Death	675 (U)	Yes	3
Bittner (2000)	Cutaneous melanoma	Death	22 (U)	Yes	6
Bohen (2003)	Non-Hodgkin's lymphoma	No response	2037 (U), 82 (S)	No	None
Devilard (2002)	Hodgkin's lymphoma	Death	1045 (U), 74 (S)	Yes	21
Dyrskjot (2003)	Bladder	Recurrence	88 (U), 26 (S)	Yes	None
Fuller (2002)	Glioma	Death	588 (U)	Yes	None
Garber (2001)	Lung adenocarcinoma	Death	366 (U)	Yes	None
Golub (1999)	AML	No response	10-50 (S)	No	15
Hofmann (2002)	ALL	Primary resistance	95 (S)	No	10
Iizuka (2003)	Hepatocellular carcinoma	Recurrence	12 (S)[a]	No	4
Kihara (2001)	Oesophageal cancer	Death	52 (S)[a]	No	8
Miura (2002)	Lung adenocarcinoma	Death	27 (S)	Yes	48
Pomeroy (2002)	Medulloblastoma	Death	8 (S), 100 (U)	Yes	23
Rosenwald (2002)	DLBCL	Death	100 (U), 17 (S)[a]	Yes	2
Shipp (2002)	DLBCL	Death	13 (S), 90 (U)	No	3
Singh (2002)	Prostate	Recurrence	5 (S)	Yes	4
Sorlie (2001)	Breast	Death	264 (S), 456 (U)[a]	Yes	2
Sotiriou (2002)	Breast	No response	37 (S)	No	7
Stratowa (2001)	B-cell CLL	Death	6 (S)	Yes	6
Takahashi (2001)	Clear cell renal cell	Death	51 (S), 3184 (U)	Yes	10
Van't Veer (2002)	Breast	Metastatic disease	70 (S), 4968 (U)[a]	Yes	24
Van de Vijver (2002)	Breast	Metastatic disease	70 (S)[a]	No	None
Virtanen (2002)	Lung	Death	4253 (U)	Yes	2
Wang (2002)	Melanoma	No complete response	4293 (U), 33 (S)	No	25
Wigle (2002)	NSCLC	Early recurrence/death	2899 (U), 22 (S)	Yes	9
Yeoh (2002)	ALL	Relapse	11322 (U), failed (S)	Yes	1
Recent studies					
Bredel (2006)	Glioblastoma	No response	10 (U), 10 (S)[a]	Yes	6
Helleman (2006)	Ovarian Cancer	No response	69 (S)[a]	No	8
Luthra (2006)	Oesophageal cancer	Metastatic disease	approximately 400 (U)	Yes	15
Nanni (2006)	Prostate	Recurrence	3000 (U), 89 (S)[a]	No	12
Pereard (2006)	Breast	Death	402 (U), 14 (S)[a]	Yes	3
Sanchez-Aguilera (2006)	Hodgkin	Death	failed (U), 145 (S)[a]	Yes	16
Sanchez-Carbayo (2006)	Bladder cancer	Death	unclear (U), 100 (S)[a]	Yes	7
Thueringen (2006)	Breast cancer	No response	512 (S)[a]	No	32
Watanabe (2006)	Rectal cancer	No response	unclear (U), 33 (S)[a]	No	5
Wei (2006)	Ovarian cancer	Relapse	112 (S)[a]	No	4
Wilson (2006)	AML	Death	9463 (U)	Yes	14
Winnepenninckx (2006)	Melanoma	Metastatic disease	11043 (U), 254 (S)[a]	Yes	10
Yamanaka (2006)	Glioma	Death	21 (S)	No	9
Zhao (2006)	Renal Cell Carcinoma	Death	3674 (U), 259 (S)[a]	Yes	3
Zirn (2006)	Wilms tumours	Death	12 (S)	Yes	47

ALL: acute lymphoblastic leukaemia; AML: acute myeloid leukaemia; CLL: chronic lymphocytic leukaemia; DLBCL: diffuse large B-cell lymphoma; NSCLC: non-small cell lung cancer; S: supervised clustering or other method of determination of molecular subtypes that is dependent on the clinical outcome; U: unsupervised clustering or other method of determination of molecular subtypes that is independent of the clinical outcome.

a Some independent validation was performed by same or different team.

Table 2 – 'Specific' comments and referencing among all mentioned genes or groups thereof

Author (year)	Total comments	Specific data or stream of reasoning (%)	Referenced (%)
Earlier studies			
Ahr (2002)	None	NA	NA
Alizadeh (2000)	None	NA	NA
Beer (2002)	9	6 (67)	6 (67)
Belbin (2002)	21	16 (76)	16 (76)
Bertucci (2002)	25	11 (44)	18 (72)
Bhattacharjee (2001)	3	3 (100)	3 (100)
Bittner (2000)	6	6 (100)	6 (100)
Bohen (2003)	None	NA	NA
Devilard (2002)	21	20 (95)	13 (62)
Dyrskjot (2003)	None	NA	NA
Fuller (2002)	None	NA	NA
Garber (2001)	None	NA	NA
Golub (1999)	15	15 (100)	7 (47)
Hofmann (2002)	10	10 (100)	3 (30)
Iizuka (2003)	4	3 (75)	3 (75)
Kihara (2001)	8	8 (100)	8 (100)
Miura (2002)	48	24 (50)	38 (79)
Pomeroy (2002)	23	5 (22)	7 (30)
Rosenwald (2002)	2	0 (0)	0 (0)
Shipp (2002)	3	3 (100)	3 (100)
Singh (2002)	4	2 (50)	2 (50)
Sorlie (2001)	2	2 (100)	2 (100)
Sotiriou (2002)	7	7 (100)	7 (100)
Stratowa (2001)	6	6 (100)	6 (100)
Takahashi (2001)	10	9 (90)	8 (80)
Van't Veer (2002)	24	24 (100)	NA
Van de Vijver (2002)	None	NA	6 (25)
Virtanen (2002)	2	2 (100)	0 (0)
Wang (2002)	25	4 (16)	4 (16)
Wigle (2002)	9	5 (56)	4 (44)
Yeoh (2002)	1	0 (0)	0 (0)
Recent studies			
Bredel (2006)	6	6 (100)	6 (100)
Helleman (2006)	8	8 (100)	8 (100)
Luthra (2006)	15	14 (93)	10 (67)
Nanni (2006)	12	8 (67)	7 (58)
Perreard (2006)	3	3 (100)	3 (100)
Sanchez-Aguilera (2006)	16	11 (69)	11 (69)
Sanchez-Carbayo (2006)	7	7 (100)	6 (86)
Thueringen (2006)	33	3 (9)	3 (9)
Watanabe (2006)	5	5 (100)	5 (100)
Wei (2006)	4	4 (100)	4 (100)
Wilson (2006)	14	11 (79)	13 (93)
Winnepenninckx (2006)	10	10 (100)	6 (60)
Yamanaka (2006)	8	8 (100)	8 (100)
Zhao (2006)	3	3 (100)	3 (100)
Zirn (2006)	47	22 (47)	22 (47)
All applicable	479	314 (65)	285 (59)

NA: Not applicable (we identified no eligible comments in these papers).

the percentage of comments that pertained to the same gene, and the percentage of comments that pertained to the same disease did not differ beyond what would be expected by chance between the early and recent studies. Given sample size limitations, we acknowledge that small or even modest differences could have been missed, but it is not likely that major differences exist.

3.4. *Transparent, publicly available data, data processing and analysis*

As shown in Table 5, six of the ten (60%) early articles (published in June 2002–February 2003) had publicly available databases, and the proportion was similar (67%, 10/15) for articles published in 2006. Nevertheless, only three studies

Table 3 – Characteristics of comments pertaining to specific data or stream of reasoning

Author (year)	Number of comments	'For' the research finding (%)	Same gene (%)	Same disease (%)
Early studies				
Ahr (2002)	NA	NA	NA	NA
Alizadeh (2000)	NA	NA	NA	NA
Beer (2002)	6	6 (100)	6 (100)	4 (67)
Belbin (2002)	16	13 (81)	12 (75)	6 (38)
Bertucci (2002)	11	5 (45)	11 (100)	6 (55)
Bhattacharjee (2001)	3	3 (100)	3 (100)	3 (100)
Bittner (2000)	6	6 (100)	6 (100)	0 (0)
Bohen (2003)	NA	NA	NA	NA
Devilard (2002)	20	20 (100)	20 (100)	12 (60)
Dyrskjot (2003)	NA	NA	NA	NA
Fuller (2002)	NA	NA	NA	NA
Garber (2001)	NA	NA	NA	NA
Golub (1999)	15	14 (93)	15 (100)	1 (7)
Hofmann (2002)	10	5 (50)	10 (100)	9 (90)
Iizuka (2003)	3	3 (100)	3 (100)	0 (0)
Kihara (2001)	8	7 (88)	8 (100)	2 (25)
Miura (2002)	24	23 (96)	7 (29)	16 (67)
Pomeroy (2002)	5	5 (100)	5 (100)	2 (40)
Rosenwald (2002)	0	0 (0)	0 (0)	0 (0)
Shipp (2002)	3	3 (100)	3 (100)	0 (0)
Singh (2002)	2	2 (100)	2 (100)	2 (100)
Sorlie (2001)	2	2 (100)	2 (100)	2 (100)
Sotiriou (2002)	7	6 (86)	7 (100)	0 (0)
Stratowa (2001)	6	5 (83)	5 (83)	3 (50)
Takahashi (2001)	9	9 (100)	9 (100)	3 (33)
Van't Veer (2002)	24	18 (75)	6 (25)	6 (25)
Van de Vijver (2002)	NA	NA	NA	NA
Virtanen (2002)	2	2 (100)	2 (100)	2 (100)
Wang (2002)	4	2 (50)	4 (100)	0 (0)
Wigle (2002)	5	5 (100)	5 (100)	0 (0)
Yeoh (2002)	0	0 (0)	0 (0)	0 (0)
Recent studies				
Bredel (2006)	6	5 (83)	6 (100)	2 (33)
Helleman (2006)	8	7 (88)	7 (88)	4 (50)
Luthra (2006)	14	14 (100)	8 (57)	8 (57)
Nanni (2006)	8	8 (100)	4 (50)	2 (25)
Perreard (2006)	3	3 (100)	3 (100)	3 (100)
Sanchez-Aguilera (2006)[a]	11	11 (100)	11(100)	0 (0)
Sanchez-Carbayo (2006)	7	7 (100)	6 (86)	7 (100)
Thueringen (2006)	3	3 (100)	3 (100)	3 (100)
Watanabe (2006)	5	5 (100)	5 (100)	1 (20)
Wei (2006)	4	4 (100)	4 (100)	1 (25)
Wilson (2006)	11	8 (73)	11(100)	11 (100)
Winnepenninckx (2006)[a]	10	10 (100)	10 (100)	0 (0)
Yamanaka (2006)	8	8 (100)	8 (100)	6 (75)
Zhao(2006)	3	0 (0)	3 (100)	3 (100)
Zirn (2006)	22	14 (64)	22 (100)	17 (77)
All applicable	314	270 (86)	262 (83)	147 (47)

Any comments 'against' the research finding were made only in Golub (1999) $n = 1$, Hofmann (2002) $n = 5$, Van't Veer (2002) $n = 6$, Bredel (2006) $n = 1$, Helleman (2006) $n = 1$, Wilson (2006) $n = 3$, Thueringen (2006) $n = 1$, Zhao (2006) $n = 3$ and Zirn (2006) $n = 8$; another 15 comments were neutral/unclear.

NA: Not applicable (we identified no eligible comments in these papers).

a In Winnepenninckx, we have not counted an online appendix that lists 70 gene tags (as opposed to just ten discussed in the full text). Even though the appendix simply lists the 70 genes, it does not specify whether they are 'for' or 'against', but the tone of the Discussion suggests that the direction is generally 'for', even though only selected genes are discussed in more detail. A similar situation arises in Sanchez-Aguilera et al.

(one early one and two published in 2006) had MIAME-compliant data for scanned images as well as raw and processed (normalised and parameterised) data (Table 5). As a post hoc analysis we explored whether these studies with better transparency of their data were also more balanced in their comments. These three studies had no clear preference for comments in favour of their gene findings: one made no comments at all, another had five 'for' and one 'against' comment and the last had three 'against' comments without any 'for' comments (overall 5/9 'for' comments as compared with

Table 4 – Comparison of the two datasets of microarrays studies

	Early studies ($n = 31$)	Recent studies ($n = 15$)	p-value
Selection method			
Unsupervised	9	2	0.30
Supervised	13	5	
Both	9	8	
Comments on genes/groups thereof			
No	7	0	0.08
Yes	24	15	
Among studies where comments were made			
Percent specific comments, median (IQR)	92 (50, 100)	100 (69, 100)	0.29
Percent referenced comments, median (IQR)	69 (30, 100)	93 (60, 100)	0.11
Among specific comments			
Percent 'for' the research finding, median (IQR)	100 (83, 100)	100 (83, 100)	0.63
Percent on the same gene, median (IQR)	100 (100, 100)	100 (88, 100)	0.68
Percent on the same disease, median (IQR)	39 (0, 66)	57 (25, 100)	0.32

IQR: interquartile range.

Table 5 – Availability of publicly accessible data, and MIAME compliance for data and array description

Author (year)	Papers refer to publicly available datasets	MIAME-compliant data	MIAME-compliant array description
Early studies			
Beer (2002)	No	No	No
Bohen (2003)	GEO (GSE3646), SMD	Images, raw, processed	Yes
Dyrskjot (2003)	GEO (GS88, GS89)	Only scaled values	Yes
Iizuka (2003)	At local website (Yamaguchi)	Raw and processed	Previously described
Miura (2002)	No	No	No
Rosenwald (2002)	At LLMPP (NIH-supported)	Raw and processed	Previously described
Van de Vijver (2002)	No	No	Previously described
Virtanen (2002)	At journal supplement	Only background-subtracted	No
Wang (2002)	No	No	No
Wigle (2002)	At local website (Toronto)	Only processed	No
Recent studies			
Bredel (2006)	SMD	Images, raw, processed	Yes
Helleman (2006)	At local website (Erasmus), but not accessible when we tried to access	Only normalised, but not accessible	Previously described?
Luthra (2006)	No	No	Yes
Nanni (2006)	GEO (GSE3868)	Only scaled values	Yes
Perreard (2006)	GEO (GSE2607), also at local website (UNC)	Raw and processed	Yes
Sanchez-Aguilera (2006)	No	No	No?
Sanchez-Carbayo (2006)	At journal supplement	Only processed	Previously described
Thueringen (2006)	GEO (GSE4056)	Raw and processed	Yes
Watanabe (2006)	GEO (GSE3493)	Only signal intensity	Yes
Wei (2006)	No	No	Previously described
Wilson (2006)	No	No	Previously described
Winnepenninckx (2006)	E_TABM_1,2,4	Raw and processed	Yes
Yamanaka (2006)	GEO (GSE4381)	Raw and processed	Yes
Zhao (2006)	GEO (GSE3538), SMD (484)	Images, raw, processed	Yes
Zirn (2006)	Array express E-MEXP-221	Only raw	Previously described

GEO: Gene expression omnibus; LLMPP: Lymphoma/leukaemia molecular profiling project; SMD: Stanford microarrays database; Previously described: array not described in public submission record, but it consists of a known array for which a MIAME-compliant public description was readily available.

265/305 in the other studies, exact $p = 0.025$). Another six studies included MIAME-compliant raw and processed datasets, although not all steps of processing were always available. Finally, seven articles referred to either raw or processed data, but not both. Array descriptions were either directly available in the public datasets or could be easily inferred in the recent studies, since commercial assays were used; this was not so for five of the ten early studies (Table 5).

The amount of detail on the image processing and normalisation and parameterisation of data differed a lot across

Table 6 – Examples of analyses with insufficient detail to fully reproduce the analysis without further communication with the original investigators; selected presentation of best analysis; and different selection choices for eligible data

Author (year)	Example
	Details need clarification
Bohen (2003)	Singular value decomposition was used to remove artifact and K-nearest neighbours impute algorithm was used to estimate missing data; references given, but use of these methods requires specification of the exact analytical options
Dyrskjot (2003)	Used average linkage clustering with a modified Pearson correlation coefficient as similarity metric (unclear which modification was used)
Virtanen (2002)	Used a lowess normalisation; reference given, but use of lowess requires specification of the exact analytical options
Nanni (2006)	Packages described without specification of parameters and options (e.g. the germa package was used for normalisation and background correction)
	Selected presentation of best analysis
Iizuka (2003)	Used support vector machine-based system that could select anywhere between the top 10 and top 300 genes; only data with top 50 genes are shown, stating that this is when the system performed best
	Different selection choices for eligible data
Bredel (2006)	Survival analysis used transcripts revealed by significance analysis of microarrays that also had expression in >75% of the specimens
Rosenwald (2002)	Selected genes where data were available for at least 90% of the patients and the gene-expression variances for the gene were in the upper 33rd percentile for such variances
Wigle (2002)	Selected 2899 genes that had data on at least 80% of the samples and of which the transcripts had at least two or more samples with an absolute value of two in log2 space
Watanabe (2006)	Samples were eligible for normalisation and further analysis when they were classified as flag-P or flag-M in >50% of the samples

studies. For feature extraction, typically the software was mentioned, but this could be accompanied by no further statement[43,50] or generic statements that feature extraction was done 'as recommended by the manufacturer'[55] or 'according to the manufacturer's instructions'.[54] There was also large variability in the details used to present normalisation procedures, thus it would not be possible to generate with certainty the same exact data processing for the studies that only provided the raw unprocessed information.

Finally, all studies that provided public access to their processed data had detailed description of the statistical analysis methods and all of them mentioned both the software and the various statistical or bioinformatics techniques that were employed in different steps of the analysis process. However, we esteem that the exact same analysis as used in each paper would practically not be possible to repeat without communication with the original investigators, given the complexity of the process and the fact that for several of the techniques the specific options employed were not fully specified (examples in Table 6). Moreover, unavoidably arbitrary selection choices had to be made for what constituted eligible data and these choices were different across studies (example in Table 6). It was uncommon to acknowledge that several analyses had been performed and only the best results were being presented (example in Table 6), but it is unknown whether this was also true of studies that did not acknowledge such selectivity.

4. Discussion

In an empirical evaluation of a large sample of gene expression profiling studies for cancer outcomes we have found that there is strong selection bias in the discussion and interpretation of the research findings. Few genes and groups thereof are selectively commented on, and supportive comments outnumber comments against the research findings by nine to one. This tendency has been similar in early studies and in studies published in 2006. Selective interpretation of the results would be less of a problem if the data from these investigations were publicly available and the data processing and analysis were fully transparent. However, we found that public availability of data has been suboptimal; even in 2006 only 10/15 studies had publicly deposited data. Only three studies had scanned images, raw and processed data available. Processing details varied across studies, but overall it would be very difficult, if not impossible, for outsider investigators to replicate the exact same analyses performed in each study.

If too many genes have the same low prognostic effect on outcomes, different studies are expected to implicate different genes each time. Thus one might expect a much higher rate of 'against' comments than supportive ones. Ein-Dor et al.[60] estimated that studies of several thousands of patients are needed to generate clinically useful prediction lists that share even 50% of the included genes in two independent studies. With limited sample sizes, if 90% of the published data on specific genes in the literature are false positives, with replication studies one would expect again most comments to be against rather than supportive. Published signatures on the prognosis of the same cancer have very low concordance and very limited or no overlap between the selected genes and this has led to the argument that probably there is exchangeability of genes (different genes having the same role in different derived signatures).[61,62] However, this is not an excuse for suboptimal analysis of the data. Consistent with our findings, Dupuy and Simon recently found that 50% of microarray studies with cancer outcomes published in 2004 had major flaws in their data analysis.[63]

For complex signatures including many genes, it may be practically impossible anyhow to make comments in the discussion for each one of these genes or even families thereof containing many members each. Selecting a few favourable

comments from a potentially vast available literature of possibly relevant information may give a false sense of corroboration of the findings by other lines of research evidence.

In discussing their findings in any field, investigators may often be tempted to show more prominently the positive aspects of their work and downplay the limitations, drawbacks and contradictory data.[64,65] Citation bias has also been demonstrated against 'unfavourable' results.[66] Selective presentation and highlighting of results is a major threat in molecular medicine,[67] given the inherent complexity and analytical flexibility of the research involved.[10]

Selection bias in the discussion of the findings apparently accompanies selection bias in data availability, processing and analysis. Despite major progress in the development of reliable public databases,[68–70] public availability and transparency remain suboptimal and discussion of microarray findings seems biased. The cancer field is probably the most advanced to-date in terms of applications of microarray technologies, and the only one where microarrays have reached the point of being approved for clinical use.[71] Cancer-related studies also represent more than half of all microarray research in humans.[1] We worry that if transparency of information, public data availability and unselected seasoned discussion are lacking in this advanced field that is moving into clinical application,[72] this may also provide a poor example for other fields.

Our observations reinforce existing recommendations that microarray data should be publicly available and the data process should be transparent and standardised.[73,74] Journals can have a major role in facilitating and enhancing public availability of information. Authors may sometimes avoid depositing information in public or may deposit less complete information, if the publishing journal does not have strict policies in this regard. We also recommend that interpretation of microarray results should avoid unilateral gene-specific claims in favour of the findings. The complete molecular signature is what matters. Finally, we recommend that if other (external) evidence is used, there should be a systematic, unbiased approach to its collection and presentation and it should include both favourable and unfavourable information. This would help place this very important research into its proper context.

Conflict of interest statement

None declared.

Appendix

Table A.1 – Operational examples for discussion of findings from microarrays studies

Characteristic	Example
Referenced comment	
Yes	ESR1 has been found to be up regulated in patients with metastatic breast cancer also in another study [reference given].
No	As above, no reference given.
Relation to research finding	
For	ESR1 is up regulated also in an experimental model of aggressive breast cancer in mice.
Against	Contrary to what we found, ESR1 is down regulated in an experimental model of aggressive breast cancer in mice.
Reasoning on exactly the same gene	
Yes	ESR1 has been found to be up regulated in patients with metastatic breast cancer also in another study
No	A different estrogen receptor gene, ESR2, has been found to be up regulated in patients with metastatic breast cancer in another study
Reasoning on exactly the same disease	
Yes	ESR1 has been found to be up regulated in patients with metastatic breast cancer also in another study.
No	ESR1 has been found up regulated also in patients with poor prognosis colorectal cancer.

Specific text comments for discussion of findings: real examples

Characteristic	Study (malignancy)	Gene and text comments (verbatim)
Referenced comment		
Yes	Helleman et al. (ovarian cancer)	ANXA4 was found to be highly expressed in clear cell compared with other histological types of ovarian carcinomas studied by Schaner et al. [reference given] and Schwartz et al. [reference given]
No	Wang et al. (melanoma)	Similarly [to other, aforementioned genes], JAK-1 and Txk, relatively suppressed in responding lesions, regulate T-cell signalling and differentiation [no reference given]
Relation to research finding		
For	Bittner et al. (malignant melanoma)	The overexpression of MK167 and TNFa [...] is in accordance with previous observations. (reference given)
Against	Hoffman et al. (acute lymphoblastic leukaemia)	*In vitro* studies showed that genomic amplification increased expression of BCR-ABL, or both and high-level expression of the multidrug resistance gene can cause resistance to STI571 in myeloid cell lines from patients with chronic myeloid leukaemia. We did not detect amplified expression of ABL.
Neutral/unclear	Wang et al. (melanoma)	Other genes are associated with transforming growth factor-b-like function (MADH3 INHBA) [unclear if the data support or contradict the finding]

(*continued on next page*)

Table A.1 – *continued*

Characteristic	Study (malignancy)	Gene and text comments (verbatim)
Reasoning on exactly the same gene		
Yes	Golub et al. (acute leukemia)	The single most highly correlated gene out of the 6817 was the homeobox gene HOXA9 which was overexpressed in patients with treatment failure. HOXA9 expression has been shown to transform myeloid cells *in vitro* and to cause leukaemia to animal models [a gene used in the molecular signature is in accordance with other studies; also a reference is given]
No	Miura et al. (lung adenocarcinoma)	For example chromosome 3p21.3 is a well known region for a frequent homologous deletion in lung cancer. Using the F test for smoking we found two genes, 101F6 and CACT, located on 3p21 region. [The genes discussed here are located on a chromosome region. This region is associated with lung cancer, but it contains dozens of genes]
Reasoning on exactly the same disease		
Yes	Devilard et al. (Hodgkin's disease)	The third gene cluster (orange cluster) underexpressed in nodular BOHD [bad outcome HD], included two tumour suppressor genes PTEN and DCC. [. . .] The PTEN deficiency observed in BOHD cases might cause inadequate responsiveness to autocrine and paracrine pathways operating in HD, thus favouring aggressive behaviour of these tumours. [Previous clinical findings referring to the same gene are in accordance to the findings of the authors]
No	Beer et al. (lung adenocarcinoma)	Elevated IGFBP3 gene expression has been found in colon cancer [comment referring to the same gene supporting the finding but in a different disease]

In the above examples, we assume that ESR1 is a gene that belongs to those identified in the molecular signature (significantly upregulated) for poor outcome in breast cancer patients, while ESR2 is not. The example is entirely hypothetical and is given to illustrate the operation of the criteria that we pursued.

REFERENCES

1. Schena M. *Microarray analysis*. Hoboken, NJ: John Wiley & Sons; 2003.
2. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003;**362**:1439–44.
3. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003;**95**:14–8.
4. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;**29**:365–71.
5. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006;**7**:55–65.
6. Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;**24**:1151–61.
7. Larsson O, Sandberg R. Lack of correct data format and comparability limits future integrative microarray research. *Nat Biotechnol* 2006;**24**:1322–3.
8. Editorial. Minimum compliance for a microarray experiment? *Nat Genet* 2006;**38**:1089.
9. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;**2**:e124.
10. Ioannidis JP. Microarrays and molecular research: noise discovery? *Lancet* 2005;**365**:454–5.
11. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;**365**:488–92.
12. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;**286**:531–7.
13. Stratowa C, Loffler G, Lichter P, et al. cDNA microarray gene expression analysis of B-cell chronic lymphocytic leukemia proposes potential new prognostic markers involved in lymphocyte trafficking. *Int J Cancer* 2001;**91**:474–80.
14. Hofmann WK, de Vos S, Elashoff D, et al. Relation between resistance of Philadelphia-chromosome-positive acute lymphoblastic leukaemia to the tyrosine kinase inhibitor STI571 and gene-expression profiles: a gene-expression study. *Lancet* 2002;**359**:481–6.
15. Miura K, Bowman ED, Simon R, et al. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res* 2002;**62**:3244–50.
16. Fuller GN, Hess KR, Rhee CH, et al. Molecular classification of human diffuse gliomas by multidimensional scaling analysis of gene expression profiles parallels morphology-based classification, correlates with survival, and reveals clinically-relevant novel glioma subsets. *Brain Pathol* 2002;**12**:108–16.
17. Wang E, Miller LD, Ohnmacht GA, et al. Prospective molecular profiling of melanoma metastases suggests classifiers of immune responsiveness. *Cancer Res* 2002;**62**:3581–6.
18. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002;**8**:68–74.
19. Takahashi M, Rhodes DR, Furge KA, et al. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc Natl Acad Sci U S A* 2001;**98**:9754–9.
20. Sotiriou C, Powles TJ, Dowsett M, et al. Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res* 2002;**4**:R3.
21. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;**8**:816–24.

22. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;**403**:503–11.
23. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;**346**:1937–47.
24. Kihara C, Tsunoda T, Tanaka T, et al. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res* 2001;**61**:6474–9.
25. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530–6.
26. Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002;**415**:436–42.
27. Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 2002;**359**:131–2.
28. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;**98**:10869–74.
29. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 2001;**98**:13790–5.
30. Belbin TJ, Singh B, Barber I, et al. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Res* 2002;**62**:1184–90.
31. Bittner M, Meltzer P, Chen Y, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000;**406**:536–40.
32. Devilard E, Bertucci F, Trempat P, et al. Gene expression profiling defines molecular subtypes of classical Hodgkin's disease. *Oncogene* 2002;**21**:3095–102.
33. Wigle DA, Jurisica I, Radulovich N, et al. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. *Cancer Res* 2002;**62**:3005–8.
34. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* 2001;**98**:13784–9.
35. Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002;**1**:203–9.
36. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002;**1**:133–43.
37. Virtanen C, Ishikawa Y, Honjoh D, et al. Integrated classification of lung tumors and cell lines by expression profiling. *Proc Natl Acad Sci U S A* 2002;**99**:12357–62.
38. Dyrskjot L, Thykjaer T, Kruhoffer M, et al. Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet* 2003;**33**:90–6.
39. Iizuka N, Oka M, Yamada-Okabe H, et al. Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *Lancet* 2003;**361**:923–9.
40. Bohen SP, Troyanskaya OG, Alter O, et al. Variation in gene expression patterns in follicular lymphoma and the response to rituximab. *Proc Natl Acad Sci U S A* 2003;**100**:1926–30.
41. Bertucci F, Nasser V, Granjeaud S, et al. Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. *Hum Mol Genet* 2002;**11**:863–72.
42. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**:1999–2009.
43. Bredel M, Bredel C, Juric D, et al. Tumor necrosis factor-alpha-induced protein 3 as a putative regulator of nuclear factor-kappaB-mediated resistance to O6-alkylating agents in human glioblastomas. *J Clin Oncol* 2006;**24**:274–87.
44. Helleman J, Jansen MP, Span PN, et al. Molecular profiling of platinum resistant ovarian cancer. *Int J Cancer* 2006;**118**:1963–71.
45. Luthra R, Wu TT, Luthra MG, et al. Gene expression profiling of localized esophageal carcinomas: association with pathologic response to preoperative chemoradiation. *J Clin Oncol* 2006;**24**:259–67.
46. Nanni S, Priolo C, Grasselli A, et al. Epithelial-restricted gene profile of primary cultures from human prostate tumors: a molecular approach to predict clinical behavior of prostate cancer. *Mol Cancer Res* 2006;**4**:79–92.
47. Perreard L, Fan C, Quackenbush JF, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res* 2006;**8**:R23.
48. Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J Clin Oncol* 2006;**24**:778–89.
49. Sanchez-Aguilera A, Montalban C, de la Cueva P, et al. Tumor microenvironment and mitotic checkpoint are key factors in the outcome of classic Hodgkin lymphoma. *Blood* 2006;**108**:662–8.
50. Thuerigen O, Schneeweiss A, Toedt G, et al. Gene expression signature predicting pathologic complete response with gemcitabine, epirubicin, and docetaxel in primary breast cancer. *J Clin Oncol* 2006;**24**:1839–45.
51. Watanabe T, Komuro Y, Kiyomatsu T, et al. Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles. *Cancer Res* 2006;**66**:3370–4.
52. Wei SH, Balch C, Paik HH, et al. Prognostic DNA methylation biomarkers in ovarian cancer. *Clin Cancer Res* 2006;**12**:2788–94.
53. Wilson CS, Davidson GS, Martin SB, et al. Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction. *Blood* 2006;**108**:685–96.
54. Winnepenninckx V, Lazar V, Michiels S, et al. Gene expression profiling of primary cutaneous melanoma and clinical outcome. *J Natl Cancer Inst* 2006;**98**:472–82.
55. Yamanaka R, Arao T, Yajima N, et al. Identification of expressed genes characterizing long-term survival in malignant glioma patients. *Oncogene* 2006;**25**:5994–6002.
56. Zhao H, Ljungberg B, Grankvist K, et al. Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS Med* 2006;**3**:e13.
57. Zirn B, Hartmann O, Samans B, et al. Expression profiling of Wilms tumors reveals new candidate genes for different clinical parameters. *Int J Cancer* 2006;**118**:1954–62.
58. Ball CA, Sherlock G, Parkinson H, et al. Standards for microarray data. *Science* 2002;**298**:539.
59. Ball CA, Sherlock G, Parkinson H, et al. The underlying principles of scientific publication. *Bioinformatics* 2002;**18**:1409.
60. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA* 2006;**103**:5923–8.
61. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;**355**:560–9.
62. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 2005;**21**:171–8.
63. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007;**99**:147–57.

64. Horton R. The hidden research paper. *JAMA* 2002;**287**:2775–8.
65. Ioannidis JPA. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol* 2007;**60**:324–9.
66. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. *J Clin Epidemiol* 2002;**55**:407–10.
67. Contopoulos-Ioannidis DG, Alexiou G, Gouvias T, Ioannidis JPA. An empirical evaluation of multifarious outcomes in pharmacogenetics: beta2 adrenoreceptor polymorphisms and asthma treatment. *Pharmacogenet Genomics* 2006;**16**:705–12.
68. Ball CA, Brazma A, Causton H, et al. Submission of microarray data to public repositories. *PLoS Biol* 2004;**2**:e317.
69. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization assay repository. *Nucleic Acids Res* 2002;**30**:207–10.
70. Brazma A, Parkinson H, Sarkans U, et al. Array Express – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 2003;**31**:68–71.
71. Abdullah-Sayani A, Bueno-de-Mesquita JM, van de Vijver MJ. Technology Insight: tuning into the genetic orchestra using microarrays—limitations of DNA microarrays in clinical practice. *Nature Clin Practice Oncol* 2006;**3**: 501–16.
72. Ioannidis JP. Is molecular profiling ready for use in clinical decision making? *Oncologist* 2007;**12**:301–11.
73. Brazma A, Krestyaninova M, Sarkans U. Standards for systems biology. *Nat Rev Genet* 2006;**7**:593–605.
74. Michiels S, Koscielny S, Hill C. Interpretation of microarray data in cancer. *Br J Cancer* 2007;**96**:1155–8.