

## Database Tool

# CancerMA: a web-based tool for automatic meta-analysis of public cancer microarray data

Julia Feichtinger<sup>1,2,\*</sup>, Ramsay J. McFarlane<sup>1,3,\*</sup> and Lee D. Larcombe<sup>4</sup>

<sup>1</sup>North West Cancer Research Fund Institute, Bangor University, Bangor, Gwynedd LL57 2UW, UK, <sup>2</sup>Institute for Genomics and Bioinformatics, Graz University of Technology, Graz, Petersgasse 14, 8010, Austria, <sup>3</sup>NISCHR Cancer Genetics Biomedical Research Unit, Bangor University, Bangor, Gwynedd LL57 2UW, UK and <sup>4</sup>Cranfield Health, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

\*Corresponding author: Tel: +44 1248 382360; Fax: +44 1248 370731; Email: r.mcfarlane@bangor.ac.uk  
Correspondence may also be addressed to Julia Feichtinger. Email: julia.feichtinger@gmail.com

Citation details: Julia Feichtinger, Ramsay J. McFarlane and Lee D. Larcombe. CancerMA: a web-based tool for automatic meta-analysis of public cancer microarray data. *Database* (2012) Vol. 2012: article ID bas055; doi:10.1093/database/bas055.

Submitted 8 September 2012; Revised 22 November 2012; Accepted 25 November 2012

The identification of novel candidate markers is a key challenge in the development of cancer therapies. This can be facilitated by putting accessible and automated approaches analysing the current wealth of 'omic'-scale data in the hands of researchers who are directly addressing biological questions. Data integration techniques and standardized, automated, high-throughput analyses are needed to manage the data available as well as to help narrow down the excessive number of target gene possibilities presented by modern databases and system-level resources. Here we present CancerMA, an online, integrated bioinformatic pipeline for automated identification of novel candidate cancer markers/targets; it operates by means of meta-analysing expression profiles of user-defined sets of biologically significant and related genes across a manually curated database of 80 publicly available cancer microarray datasets covering 13 cancer types. A simple-to-use web interface allows bioinformaticians and non-bioinformaticians alike to initiate new analyses as well as to view and retrieve the meta-analysis results. The functionality of CancerMA is shown by means of two validation datasets.

Database URL: <http://www.cancerma.org.uk>

## Introduction

Cancer is a multi-factorial disease that can arise from alterations in expression levels of oncogenes and tumour suppressor genes, details of which can be elucidated by means of expression data (1). In the last decade, a large amount of microarray data for gene expression profiles has become available in public repositories such as ArrayExpress (2) and Gene Expression Omnibus (GEO) (3), which provide the opportunity to retrieve, reanalyse and integrate the data (4). Retrieval and reanalysis of publicly available data allow the development of automated pipelines to ensure a broad spectrum of users can execute rapid, homogeneous and reproducible analyses across a large number of datasets, addressing novel and specific questions. Data integration techniques, so-called meta-analyses, aim to combine

the data available and integrate information from multiple independent but related microarray studies to identify significant genes [reviewed by Feichtinger *et al.* (5)]. Combining studies can enhance reliability and generalizability of the results (6) and can be used to obtain a more precise estimate of gene expression. In particular, the benefit of enhancing the statistical power can help to overcome the most profound limitation of microarray studies: testing tens of thousands of hypotheses, relying only on a relatively low number of samples (7, 8). For example, Arasappan *et al.* (9) found a refined expression signature for systemic lupus erythematosus, and Vierlinger *et al.* (10) reported the identification of a potential biomarker for papillary thyroid carcinoma by means of meta-analysis approaches.

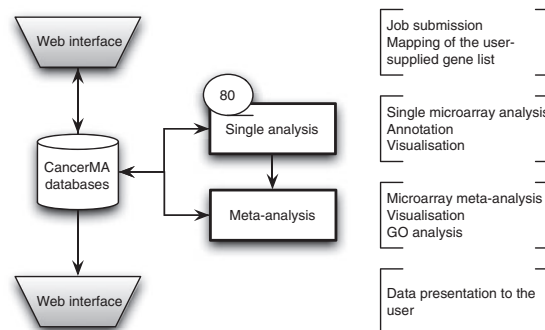
Here we present CancerMA, an openly accessible integrated bioinformatic analytical pipeline with a user-friendly and intuitive web interface to automate the reanalysis of public cancer microarray datasets with user-defined sets of biologically significant and related genes. The underlying analytical approach was developed for a previous study to identify a cohort of novel cancer-specific marker genes (11) and was automated forming the core of the CancerMA tool. Further analyses and visualizations were added to aid the data interpretation. This tool allows bioinformaticians and non-bioinformaticians alike, to obtain refined and integrated differential expression for their genes of interest across a manually curated database of 80 datasets and 13 cancer types as well as to investigate the relationships between cancer types and to reveal commonalities among them. Furthermore, it can help to narrow down the excessive number of target gene possibilities presented by modern databases and system-level resources to a manageable number of putative candidates, which can be followed up in the laboratory and/or fed into an interaction network analysis. Thus, it puts a meta-analysis pipeline in the hands of those asking the biological questions. To validate our approach, we have analysed two experimentally derived datasets from the literature and could reproduce the published results.

## Methods and structure of CancerMA

CancerMA consists of a web interface, a set of pipelined analyses and two relational databases, one holding the analysis data for each user and another one holding the gene annotation data. The general workflow is visualized in Figure 1.

### Cancer dataset retrieval

We searched for raw data of patient-derived, untreated cancer samples with corresponding normal samples deposited in the ArrayExpress (2) or the GEO (3) repository using the HG-U133 Plus 2 array from Affymetrix. After manual assessment, we divided the retrieved datasets according to the cancer type, subtype and stage. We omitted datasets with less than three control or cancer samples as well as datasets deriving from foetal tissues, tissues influenced by other diseases or cancer-associated tissues (e.g. tumour microenvironment). We could obtain 92 datasets from 50 experiments covering 13 distinct cancer types. To allow a meta-analysis, at least two datasets per cancer type were required. Subsequently, quality control using the 'simpleaffy' R package (12) was used to further assess the datasets. Based on the guidelines from Affymetrix/'simpleaffy' (available at: <http://www.bioconductor.org/packages/release/bioc/vignettes/simpleaffy/inst/doc/QCandSimpleaffy>).



**Figure 1.** CancerMA workflow. The web interface box indicates the areas where the user provides input and/or can view the mapping or analysis results. The analysis is carried out automatically without any user input. The single analysis determines the differential expression for 80 cancer microarray datasets individually, whereas the meta-analysis combines the results from the individual analyses to a differential meta-expression profile.

pdf), datasets with scale factors with 3-fold of one another, an ActB 3':mid ratio  $<3$  and a GAPDH 3':mid ratio  $<1.25$  were selected. Scale factors assess the comparability of the arrays, whereas the signal ratios of ActB and GAPDH can be used to measure the RNA quality. Based on this assessment, we omitted 12 datasets and excluded individual CEL files of 37 datasets. Finally, 80 individual curated cancer datasets originating from 45 experiments and covering 13 different cancer types (Supplementary Table S1) remained. For more details, refer to Feichtinger *et al.* (11). The full list of 80 datasets, including the GEO/ArrayExpress accession numbers as well as the 13 cancer types covered, are available on the CancerMA website (<http://www.cancerma.org.uk/information.html>). Additional experimental datasets can be obtained from the microarray repositories and added to the pipeline by the authors as they become available.

### The CancerMA pipeline and databases

The pipeline handles the single microarray analysis, the meta-analysis, the GO analysis as well as the annotation and the visualizations.

After manual assessment and quality control, all 80 datasets described above were individually pre-processed (background correction, normalization and computation of expression values) according to methods described by Hubbell *et al.* (13) using the 'affy' R package from Bioconductor (14), which assures uniformity of the analysis process.

For gene and probe annotation purposes, the Ensembl database (15), the HUGO Gene Nomenclature Committee (HGNC) database (16) and the annotation files provided by Affymetrix (available at: <http://www.affymetrix.com/support/technical/annotationfilesmain.affx>) were established as a local MySQL database.

When a new job is submitted, the user-supplied gene list is used to filter the 80 pre-processed datasets in order to reduce the number of features and enhance the statistical power (17). The 'Limma' R package (18) from Bioconductor is used to compute differentially expressed genes, and the resulting *P*-values are adjusted for multiple testing with Benjamini and Hochberg's method to control the false discovery rate (19). For the single array analysis, genes with a *P*-value <0.05 and a log<sub>2</sub>-fold change >1 are selected as potentially significant.

Subsequently, the results of the 80 individual analyses are combined. A meta-*P*-value and a meta-log<sub>2</sub>-fold change value are calculated for each cancer type (Supplementary Table S1) as well as for all cancers in total using Stouffer's method (20) and weighted linear combination (21), respectively. As two-sided *P*-values are oblivious to the effect direction, these *P*-values need to be converted to the corresponding one-sided *P*-values for up- and down-regulation separately (22). In case of multiple probes mapping to the same gene identifier, the most extreme log<sub>2</sub>-fold change value with its corresponding *P*-value are further used for feature selection. Genes with a |meta-log<sub>2</sub>-fold change| >1 or a confidence interval that does not span 0, and a meta-*P*-value <0.05 are considered as potentially significant.

Finally, all significantly up- and downregulated genes of the meta-analysis are fed into a gene ontology (GO) enrichment analysis using the 'GOstats' R package (23) from Bioconductor.

To visualize the analysis results, Circos plots (24), forest plots (25) and Krona plots (26) are created. All data belonging to a user are stored for 30 days in the CancerMA user database, which can be accessed using the web interface during this time. This analytical approach was developed for a previous study published by the authors, and automated for the basis of the CancerMA tool. For more details, refer to Feichtinger *et al.* (11).

### The CancerMA web interface

First, the CancerMA web interface handles the mapping of a user-supplied gene list as well as the subsequent job submission. Second, it allows the user to access the analysis results.

When submitting a new job, the user supplies a list consisting either of Ensembl IDs or of gene names, for which the identifiers are then mapped to their appropriate Affymetrix IDs by the tool to tell the user which genes can be analysed. Finally, the job can be submitted by providing an email address.

When viewing a finished job, the results of the various analyses and the visualizations are presented to the user in a simple-to-use web interface. All result files are also available for download. To view an example, visit <http://www.cancerma.org.uk>.

### Implementation

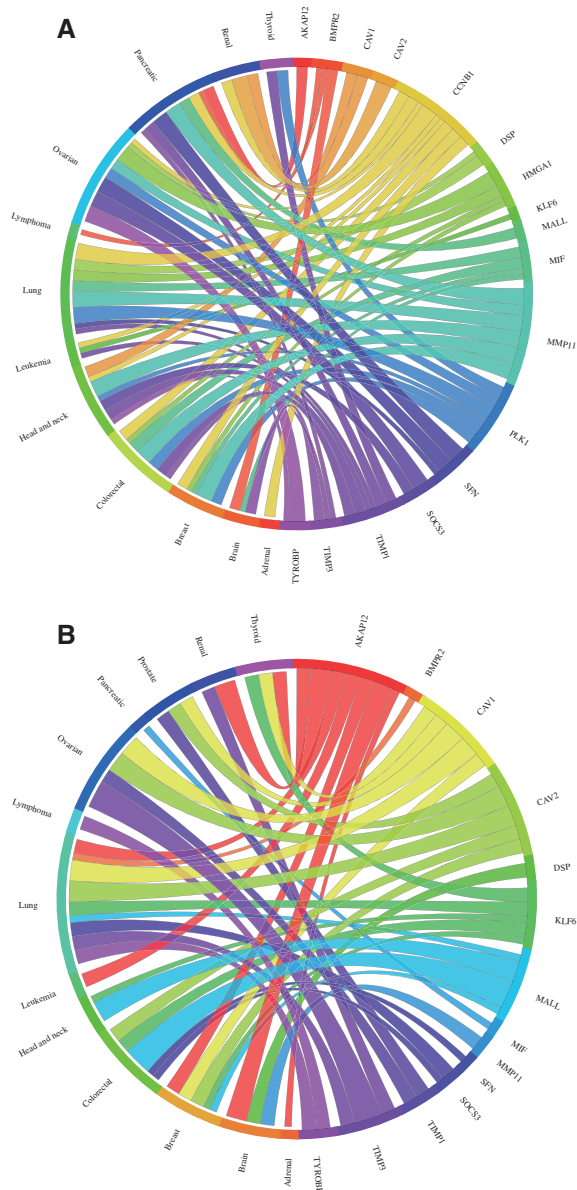
CancerMA is running on an Intel core i7 2.66 Ghz workstation with 12 Gb RAM and installed with CentOS 5.4 GNU Linux OS (x86\_64). For the relational databases, MySQL 5.0.77 (available at: <http://www.mysql.com>) was used. The CancerMA web interface was implemented using: HTML/CSS, Twitter Bootstrapp (available at: <http://twitter.github.com/bootstrap/>), Javascript/jQuery (available at: <http://jquery.com/>) and Perl 5.8.8 (available at: <http://www.perl.org>). The CancerMA pipeline was implemented using: R 2.12.1 (available at: <http://www.cran.r-project.org>) (27); the Bioconductor package (available at: <http://www.bioconductor.org>) (28) and Perl 5.8.8 (available at: <http://www.perl.org>). CancerMA is freely available online at <http://www.cancerma.org.uk>.

### Use of CancerMA

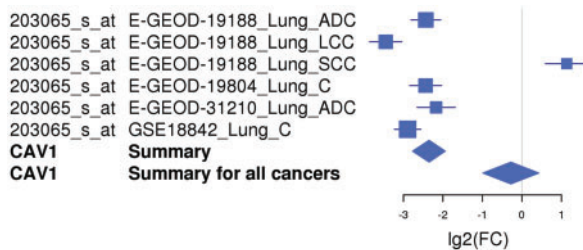
CancerMA was developed for automated computation of the differential meta-expression for genes of interest to biologists/clinicians and, in particular, as a user-friendly and intuitive tool to view and interpret the analysis results. The CancerMA web interface for viewing the analysis data consists of three sections: a general overview, the information section as well as the result section. The general overview provides basic information about the submitted job and the data available to the user. The information section includes among others the annotated genes of interest and information about the datasets used in the analysis. The result section includes the analysis results of the meta-analysis, of the single analyses, of the single analyses (only) and of the GO enrichment analysis. The meta-analysis results comprise tables with statistical values and visualizations for the meta-upregulated as well as for the meta-downregulated genes of interest. The GO analysis results contain the enriched GO terms for the meta-up- and the meta-downregulated genes, respectively. The single analysis results show all up- and downregulations of the genes of interest in all individually analysed datasets. The single analysis (only) results, however, provide genes of interest which are either consistently up- or downregulated across the datasets. Circos and Krona plots visualize the single and meta-analysis results in their entirety to highlight relationships within the data. Furthermore, forest plots visualize the meta-analysis results for each gene separately. For a detailed documentation, please refer to the CancerMA help section (<http://www.cancerma.org.uk/help.html>).

### Validation

We used two experimentally determined datasets providing genes differentially expressed in cancer to validate our analysis results and demonstrate the utility and the functionality of the tool: (i) 10 upregulated and 9 downregulated genes in lung cancer determined by cDNA array



**Figure 2.** Circos plots showing the meta-change in gene expression in relation to corresponding cancer types. The plot shows the meta-up- and meta-downregulated genes of the validation dataset from Kettunen *et al.* (29): (A) The expression of the genes *DSP*, *CCNB1*, *PLK1*, *MIF*, *HMGA1*, *SFN*, *TIMP1* and *MMP11* was found to be upregulated, whereas (B) the expression of the genes *AKAP12*, *BMPR2*, *COPEB/KLF6*, *SOCS3*, *BENE/MALL*, *TIMP3*, *CAV1*, *CAV2* and *TYROBP* was found to be downregulated in lung cancer consistent with the published results. Each connection between a gene and a cancer type indicates a statistically significant mean up- or downregulation for that cancer type derived from a number of combined array studies for cancer tissue versus normal tissue. The weight of the connection corresponds to the magnitude of the meta-change in gene expression.

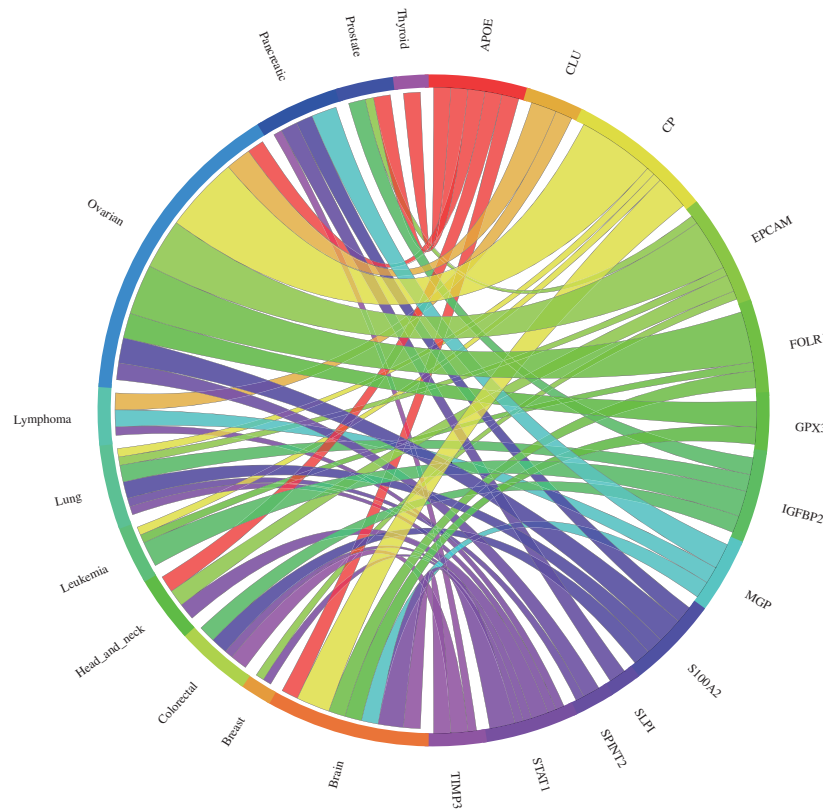


**Figure 3.** An example of a forest plot showing the expression of gene *CAV1* downregulated in lung cancer. The expression of the *CAV1* gene is downregulated in five of six microarray studies and upregulated in one study. The forest plot shows the meta-log 2-fold change values for the individual studies as well as the total values for lung cancer and for all cancer types combined. Each study is illustrated by a square; the position on the x-axis representing the measure estimate (log2FC ratio), the size proportional to the weight of the study and the horizontal line through it reflecting the confidence interval of the estimate.

analysis and partially validated by RT-PCR (29) and (ii) 13 upregulated genes in ovarian cancer validated by RT-PCR (30).

The meta-analysis results of the 17 differentially expressed genes in lung cancer (two genes reported to be upregulated were not present on the arrays used by CancerMA) were consistent with the findings described by Kettunen *et al.* (29) (Figure 2). Most genes determined to be up- or downregulated in this study were reported in various other publications to be up- or downregulated accordingly (31–42). For example, the expression of the gene *CAV1* was found to be highly downregulated in five of six cancer microarray datasets (Figure 3). This also provides a good example for the capability of meta-analysis techniques to identify a more valid set of differentially expressed genes, as biological, experimental and technological variations, including differences in experimental conditions, tissues, cell lines, species, platforms, sample treatment and processing can lead to inconsistencies in gene expression, which reflect the differences in the experimental setting in addition to the objective studied (43). Furthermore, interesting patterns emerge from our meta-analysis results; for example, the expression of *PLK2*, *MMP11*, *CCNB1* and *TIMP1* is mainly upregulated in cancer (Figure 2A), whereas the expression of *AKAP12*, *CAV1*, *CAV2*, *COPEB/KLF6* and *BENE/MALL* is mainly downregulated in cancer (Figure 2B). Additionally, commonalities between cancer types can be inferred; for example, the expression pattern found in lung cancer is highly similar to the one in colorectal, ovarian and breast cancer, in particular for the upregulated genes (Figure 2A).

Our meta-analysis of the ovarian cancer validation dataset resulted in eight genes significantly upregulated, consistent with the results described by Hough *et al.* (30), four genes not differentially expressed and one gene



**Figure 4.** Circos plot showing the meta-change in gene expression in relation to corresponding cancer types. The plot shows the meta-upregulated genes of the validation dataset from Hough *et al.* (30): The expression of the genes *GPX3*, *CLU*, *EPCAM*, *SPINT2*, *FOLR1*, *S100A2*, *APOE* and *CP* was found to be upregulated in ovarian cancer consistent with the published results. Each connection between a gene and a cancer type indicates a statistically significant mean up- or downregulation for that cancer type derived from a number of combined array studies for cancer tissue versus normal tissue. The weight of the connection corresponds to the magnitude of the meta-change in gene expression.

downregulated (Figure 4). Almost all upregulated genes were reported in various other publications to also be upregulated in ovarian cancers (44–51). Several clinical trials examining the efficiency of an immunotherapy targeting the products of these genes are currently running (52, 53). According to our meta-analysis, *TIMP3* expression was found to be significantly downregulated in ovarian cancer. However, this is consistent with the findings that *TIMP3* is a possible tumour suppressor gene. An analysis of DNA copy number and gene expression of 22q in 18 ovarian carcinomas has shown that copy number loss across the *TIMP3* locus is frequent, leading to decreased detectable *TIMP3* mRNA levels (54). Furthermore, *TIMP3* expression was reported to be downregulated in the lung cancer validation dataset that we used (29) and Hough *et al.* (30) noted that *TIMP3* was not highly or consistently expressed in their tumour samples. The four genes (*IGFBP2*, *MGP*, *STAT1* and *SLP1*) not showing significant upregulation appear to lack consistency in expression across tumour samples and/or cancer subtypes, as according to the single microarray analysis they are upregulated just in some microarray datasets (Supplementary Figure S1). This is

also consistent with the findings of Hough *et al.* (30, 50), as they report that *IGFBP2* was not consistently expressed between their tumour samples. Furthermore, *STAT1* was reported to be overexpressed only in certain subtypes of serous ovarian carcinomas (55).

#### Example workflow

In our previously published work (11) we have analysed human meiotic genes using the analytical approach now implemented into CancerMA and, with RT-PCR experimental validation, identified a novel, clinically relevant subgroup of the cancer/testis gene family (the meiCT genes), which have potential as novel cancer markers and therapeutic targets. This work serves as an example workflow for potential users.

## Discussion

### Purposes and benefits of CancerMA

CancerMA allows the automated computation of the differential meta-expression for genes of interest to biologists/clinicians across 80 cancer microarray-derived

datasets covering 13 cancer types. As shown by the validation, our meta-analysis approach enhances the statistical power by increasing the sample size and can resolve conflicting conclusions between individual studies by finding a more valid set of differentially expressed genes. Furthermore, our pipeline approach focuses on the meta-analysis on a set of related genes specified by the user, which additionally serves to enhance the significance and accuracy of the analysis, and also to narrow down the excessive number of possibilities presented by whole genome arrays to a manageable number of putative leads (17). Direct experimental evidence or other inferred relationships, such as genes involved in interaction networks, can serve as a basis to compile a set of related genes. Relationships within a gene set could include co-expression, co-regulation, affiliation to the same pathway or biological process as well as common pathological involvement.

Screening for the differential expression within a given set of genes could reveal diagnostic, therapeutic and prognostic strategies and applications for specific cancer types as well as uncover common dysfunction of specific genes, gene modules or pathways across various cancer types. Furthermore, genome-scale meta-analysis can reveal common drivers of change or similar expression modules across various cancer types and therefore point towards conserved disrupted pathways or mechanisms in cancer; for example, the p53 pathway is often disrupted in cancer either due to point mutations in *TP53* gene or due to one of the numerous alternative gene mutations that may lead to disruption of this pathway at key points [reviewed by Vogelstein *et al.* (56)]. Genetic alterations in different genes can often manifest a similar or common phenotype where these genes are related as part of the same pathway. The fact that mutations in a vast number of genes have been associated with cancer, yet disruption of only a few key pathways may give rise to the characteristics of cancer, highlights the importance of focussing on sets of related or interacting genes [reviewed by Vogelstein and Kinzler (1)].

CancerMA relies on the availability of public microarray data. Currently, we can cover 13 cancer types, but we hope that further datasets will become available in due course allowing us to expand the meta-analysis. Furthermore, we have selected datasets using the Affymetrix UG-133 Plus 2 array, as this array type is widely used and covers a large proportion of the human genome. Nevertheless, a number of genes (in particular, novel gene discoveries) are not covered by this array type and thus cannot be evaluated by this tool. However, we intend to continue the development of this tool, extending CancerMA to incorporate other Affymetrix array types and arrays from other platforms such as Illumina in due course.

### Comparison to databases and tools currently available

Additionally to the repositories storing microarray data such as ArrayExpress and GEO (3), more specialized databases have become available; for example, databases such as M<sup>2</sup>DB (57) and M<sup>3D</sup> (58) collected microarray data and uniformly pre-processed it, but do not provide data analysis and integration. Web platforms such as Oncomine (59), GEO Profiles (3), Gene Expression Atlas (available at: <http://www.ebi.ac.uk/gxa/>) or Gemma (60) focus on gene expression profiles across multiple conditions and tissues but do not combine the results of the individual studies. Web platforms such as GeneSapiens (61) and Genevestigator (62) combine individual studies by pooling and subsequently analysing the data with traditional techniques but do not use meta-analysis approaches. However, various microarray meta-analysis approaches are available as R packages such as metaMA (63), Rankprod (64) and metaArray (65), but require skills in statistics and R. Therefore, a simple-to-use web tool such as CancerMA providing the computation of the meta-expression profile using manually curated, patient-derived cancer microarrays for a set of genes of interests to biologists/clinicians to a wide audience is not yet available to our knowledge [for a detailed review of meta-analysis databases and tools, refer to Feichtinger *et al.* (5)].

## Conclusion

In summary, we present CancerMA, an integrated bioinformatic analytical pipeline to automate the identification of novel candidate cancer markers/targets by means of analysing the expression of user-supplied gene lists across a manually curated database of 80 publicly available cancer microarray datasets and 13 cancer types. Such a meta-analysis enhances reliability and generalizability of the analysis results and leads to a more precise estimate of gene expression. Furthermore, the pipeline facilitates automated, homogeneous and reproducible analysis across a large number of datasets, and establishing a simple-to-use online web interface to access the pipeline puts specialist meta-analyses in the hands of biologists.

## Supplementary Data

Supplementary data are available at *Database* online.

## Acknowledgements

The authors would like to thank Dr Sarah Morgan and Dr Jane Wakeman for their assistance in proofreading the manuscript.

## Funding

The Welsh National Institute for Social Care and Health Research (NISCHR) (HS/09/008 to J.F.); North West Cancer Research Fund grant (NWCRF) (CR888 to R.J.M.). Funding for open access charge: NISCHR (award number: HS/09/008).

*Conflict of interest.* None declared.

## References

- Vogelstein,B. and Kinzler,K.W. (2004) Cancer genes and the pathways they control. *Nat. Med.*, **10**, 789–799.
- Parkinson,H., Sarkans,U., Kolesnikov,N. et al. (2011) ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **39**, D1002–D1004.
- Barrett,T., Troup,D.B., Wilhite,S.E. et al. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
- Moreau,Y., Aerts,S., De Moor,B. et al. (2003) Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet.*, **19**, 570–577.
- Feichtinger,J., Thallinger,G.G., McFarlane,R.J. et al. (2012) Microarray meta-analysis: From data to expression to biological relationships. In: Trajanoski,Z. (ed). *Computational Medicine*. Springer, New York, NY, pp. 59–77.
- Ramasamy,A., Mondry,A., Holmes,C.C. et al. (2008) Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.*, **5**, 13.
- Campaign,A. and Yang,Y.H. (2010) Comparison study of microarray meta-analysis methods. *BMC Bioinformatics*, **11**, 408.
- Normand,S.L. (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Stat. Med.*, **18**, 321–359.
- Arasappan,D., Tong,W., Mummaneni,P. et al. (2011) Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. *BMC Med.*, **9**, 65.
- Vierlinger,K., Mansfeld,M.H., Koperek,O. et al. (2011) Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and quantification of its discriminatory power in independent validation. *BMC Med. Genomics*, **4**, 30.
- Feichtinger,J., Aldeaij,J., Anderson,R. et al. (2012) Meta-analysis of clinical data using human meiotic genes identifies a novel cohort of highly restricted cancer-specific marker genes. *Oncotarget*, **3**, 843–53.
- Wilson,C.L. and Miller,C.J. (2005) Simpleaffy: A BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, **21**, 3683–3685.
- Hubbell,E., Liu,W.-M. and Mei,R. (2002) Robust estimators for expression analysis. *Bioinformatics*, **18**, 1585–1592.
- Gautier,L., Cope,L., Bolstad,B.M. et al. (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
- Flicek,P., Amode,M.R., Barrell,D. et al. (2011) Ensembl 2012. *Nucleic Acids Res.*, **40**, 84–90.
- Seal,R.L., Gordon,S.M., Lush,M.J. et al. (2011) genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.
- Scholten,D. and von Heydebreck,A. (2005) Analysis of differential gene expression studies. In: Gentleman,R., Carey,V., Huber,W. et al. (eds), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, pp. 229–248.
- Smyth,G.K. (2005) Limma: Linear models for microarray data. In: Gentleman,R., Carey,V., Huber,W. et al. (eds), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, NY, pp. 397–420.
- Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **57**, 289–300.
- Stouffer,S.A. (1949) *The American Soldier. Studies in Social Psychology in World War II*. Princeton University Press, Princeton, NJ.
- Morgan,A.A., Khatri,P., Jones,R.H. et al. (2010) Comparison of multiplex meta analysis techniques for understanding the acute rejection of solid organ transplants. *BMC Bioinformatics*, **11**, S6.
- Zaykin,D.V. (2011) Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.*, **24**, 1836–1841.
- Falcon,S. and Gentleman,R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23**, 257–258.
- Krzywinski,M., Schein,J., Birol,I. et al. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
- Lewis,S. and Clarke,M. (2001) Forest plots: trying to see the wood and the trees. *BMJ*, **322**, 1479–1480.
- Ondov,B.D., Bergman,N.H. and Phillippy,A.M. (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinf.*, **12**, 385.
- R Development Core Team,R. (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, p. 409.
- Gentleman,R.C., Carey,V.J., Bates,D.M. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
- Kettunen,E., Anttila,S., Seppänen,J.K. et al. (2004) Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. *Cancer Genet. Cytogenet.*, **149**, 98–106.
- Hough,C.D., Cho,K.R., Zonderman,A.B. et al. (2001) Coordinately up-regulated genes in ovarian cancer. *Cancer Res.*, **61**, 3869–3876.
- Boelens,M.C., Van Den Berg,A., Vogelzang,I. et al. (2007) Differential expression and distribution of epithelial adhesion molecules in non-small cell lung cancer and normal bronchus. *J. Clin. Pathol.*, **60**, 608–614.
- Soria,J.C., Jang,S.J., Khuri,F.R. et al. (2000) Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. *Cancer Res.*, **60**, 4000–4004.
- Wang,Z.-X., Xue,D., Liu,Z.-L. et al. (2012) Overexpression of polo-like kinase 1 and its clinical significance in human non-small cell lung cancer. *Int. J. Biochem. Cell Biol.*, **44**, 200–210.
- Campa,M.J., Wang,M.Z., Howard,B. et al. (2003) Protein expression profiling identifies macrophage migration inhibitory factor and cyclophilin A as potential molecular targets in non-small cell lung cancer. *Cancer Res.*, **63**, 1652–1656.
- Hillion,J., Wood,L.J., Mukherjee,M. et al. (2009) Upregulation of MMP-2 by HMGA1 promotes transformation in undifferentiated, large-cell lung cancer. *Mol. Cancer Res.*, **7**, 1803–1812.
- Aljada,I.S., Ramnath,N., Donohue,K. et al. (2004) Upregulation of the tissue inhibitor of metalloproteinase-1 protein is associated with progression of human non-small-cell lung cancer. *J. Clin. Oncol.*, **22**, 3218–3229.

37. Coussens,L.M., Fingleton,B. and Matrisian,L.M. (2002) Matrix metalloproteinase inhibitors and cancer: trials and tribulations. *Science*, **295**, 2387–2392.
38. Jo,U., Whang,Y.M., Kim,H.K. et al. (2006) AKAP12 $\alpha$  is associated with promoter methylation in lung cancer. *Cancer Res. Treat.*, **38**, 144–151.
39. Ito,G., Uchiyama,M., Kondo,M. et al. (2004) Krüppel-like factor 6 is frequently down-regulated and induces apoptosis in non-small cell lung cancer cells. *Cancer Res.*, **64**, 3838–3843.
40. He,B., You,L., Uematsu,K. et al. (2003) SOCS-3 is frequently silenced by hypermethylation and suppresses cell growth in human lung cancer. *Proc. Natl Acad. Sci. USA*, **100**, 14133–14138.
41. Garofalo,M., Di Leva,G., Romano,G. et al. (2009) miR-221&222 regulate TRAIL resistance and enhance tumorigenicity through PTEN and TIMP3 downregulation. *Cancer Cell*, **16**, 498–509.
42. Wikman,H., Seppänen,J.K., Sarhadi,V.K. et al. (2004) Caveolins as tumour markers in lung cancer detected by combined use of cDNA and tissue microarrays. *J. Pathol.*, **203**, 584–593.
43. Cahan,P., Rovegno,F., Mooney,D. et al. (2007) Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. *Gene*, **401**, 12–18.
44. Lee,H.J., Do,J.H., Bae,S. et al. (2010) Immunohistochemical evidence for the over-expression of Glutathione peroxidase 3 in clear cell type ovarian adenocarcinoma. *Med. Oncol.*, **28**, S522–S527.
45. Xie,D., Lau,S.H., Sham,J.S.T. et al. (2005) Up-regulated expression of cytoplasmic clusterin in human ovarian carcinoma. *Cancer*, **103**, 277–283.
46. Kim,J.-H., Herlyn,D., Wong,K. et al. (2003) Identification of epithelial cell adhesion molecule autoantibody in patients with ovarian cancer. *Clin. Cancer Res.*, **9**, 4782–4791.
47. Foulkes,W.D., Campbell,I.G., Stamp,G.W. et al. (1993) Loss of heterozygosity and amplification on chromosome 11q in human ovarian cancer. *Br. J. Cancer*, **67**, 268–273.
48. Santin,A.D., Zhan,F., Bellone,S. et al. (2004) Gene expression profiles in primary ovarian serous papillary tumors and normal ovarian epithelium: identification of candidate molecular markers for ovarian cancer diagnosis and therapy. *Int. J. Cancer*, **112**, 14–25.
49. Chen,Y.-C., Pohl,G., Wang,T.-L. et al. (2005) Apolipoprotein E is required for cell proliferation and survival in ovarian cancer. *Cancer Res.*, **65**, 11424–11431.
50. Hough,C.D., Sherman-Baust,C.A., Pizer,E.S. et al. (2000) Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. *Cancer Res.*, **60**, 6281–6287.
51. Lee,C.M., Lo,H.-W., Shao,R.-P. et al. (2004) Selective activation of ceruloplasmin promoter in ovarian tumors: potential use for gene therapy. *Cancer Res.*, **64**, 1788–1793.
52. Baeuerle,P.A. and Gires,O. (2007) EpCAM (CD326) finding its role in cancer. *Br. J. Cancer*, **96**, 417–423.
53. Fisher,R.E., Siegel,B.A., Edell,S.L. et al. (2008) Exploratory study of 99mTc-EC20 imaging for identifying patients with folate receptor-positive solid tumors. *J. Nucl. Med.*, **49**, 899–906.
54. Benetkiewicz,M., Wang,Y., Schaner,M. et al. (2005) High-resolution gene copy number and expression profiling of human chromosome 22 in ovarian carcinomas. *Genes, Chromosomes Cancer*, **42**, 228–237.
55. Meinhold-Heerlein,I., Bauerschlag,D., Hilpert,F. et al. (2005) Molecular and prognostic distinction between serous ovarian carcinomas of varying grade and malignant potential. *Oncogene*, **24**, 1053–1065.
56. Vogelstein,B., Lane,D. and Levine,A.J. (2000) Surfing the p53 network. *Nature*, **408**, 307–310.
57. Cheng,W.-C., Tsai,M.-L., Chang,C.-W. et al. (2010) Microarray meta-analysis database (M2DB): a uniformly pre-processed, quality controlled, and manually curated human clinical microarray database. *BMC Bioinformatics*, **11**, 421.
58. Faith,J.J., Driscoll,M.E., Fusaro,V.A. et al. (2008) Many microbe microarrays database: Uniformly normalized affymetrix compendia with structured experimental metadata. *Nucleic Acids Res.*, **36**, D866–D870.
59. Rhodes,D.R., Kalyana-Sundaram,S., Mahavisno,V. et al. (2007) Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*, **9**, 166–180.
60. Zoubarov,A., Hamer,K.M., Keshav,K.D. et al. (2012) Gemma: A resource for the re-use, sharing and meta-analysis of expression profiling data. *Bioinformatics*, **28**, 2272–2273.
61. Kilpinen,S., Autio,R., Ojala,K. et al. (2008) Systematic bioinformatic analysis of expression levels of 17,330 human genes across 9,783 samples from 175 types of healthy and pathological tissues. *Genome Biol.*, **9**, R139.
62. Hruz,T., Laule,O., Szabo,G. et al. (2008) Genevestigator V3: A reference expression database for the meta-analysis of transcripts. *Adv. Bioinformatics*, **2008**, 420747.
63. Marot,G., Foulley,J.-L., Mayer,C.-D. et al. (2009) Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics*, **25**, 2692–2699.
64. Hong,F., Breitling,R., McEntee,C.W. et al. (2006) RankProd: a bio-conductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*, **22**, 2825–2827.
65. Choi,H., Shen,R., Chinnaiyan,A.M. et al. (2007) A latent variable approach for meta-analysis of gene expression data from multiple microarray experiments. *BMC Bioinformatics*, **8**, 364.