

Brief communication

Gene expression profiles that segregate patients with childhood acute lymphoblastic leukaemia: An independent validation study identifies that endoglin associates with patient outcome

Daniel Catchpoole^{a,*}, Andy Lail^a, Dachuan Guo^a, Qing-Rong Chen^b, Javed Khan^b

^a *The Tumour Bank, The Oncology Research Unit, The Children's Hospital at Westmead, Locked Bag 4001, Westmead, NSW 2145, Australia*

^b *Oncogenomics Section, Pediatric Oncology Branch, National Cancer Institute, National Institute of Health, Advanced Technology Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA*

Received 26 October 2006; received in revised form 26 April 2007; accepted 29 April 2007

Available online 18 June 2007

Abstract

In this report, we determine whether genes identified in a previously reported cDNA microarray investigation of childhood acute lymphoblastic leukaemia (ALL) diagnostic bone marrow have the same distinguishing power in an independently derived cDNA microarray dataset from an equivalent but distinct patient cohort. Genes previously reported as discriminatory, generally were unable to distinguish ALL lymphocyte lineages, the presence of the Tel-AML1 translocation and patient risk stratification. An artificial neural network identified *endoglin*, which was reported in the initial study as a potential lineage marker, was actually better at identifying ALL patients with poor outcome.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Acute lymphoblastic leukaemia; cDNA microarray; Gene expression profiles; Validation; Endoglin

1. Introduction

There have been numerous investigations into the classification of childhood acute lymphoblastic leukaemia (ALL) using microarray generated gene expression profiles. The ultimate intention of a number of these investigations is to identify sets of genes whose expression correlate for disease progression, clinical outcome or the identification of novel subtypes of ALL. In general, the approach has been to interrogate the microarray gene expression data for simplified common 'signatures' which consist of a reduced numbers of genes which best discriminate ALL patient subgroups. Critical to robust research is the ability to replicate experimental findings. Within the field of microarray analysis, validation of the expression of identified 'significant genes' is required to be undertaken in independent laboratories on equivalent but separately derived data sets.

The study by Moos et al. [1] reported the analysis of 51 bone marrow samples from children with acute leukaemia using cDNA spotted microarray. Gene expression data from 4608 genes in leukaemic blasts were assessed using parametric *t*-test comparison (two tailed, homoscedastic) and Infocore, a non-parametric rank-based scoring system formulated on conditional entropy ([1] and references therein). Using these two supervised methods the investigators identified a signature of 20 discriminating genes in an attempt to demonstrate how microarray gene expression profiling may augment current risk-based classification of paediatric leukaemia [1].

In this paper, we report the examination of the genes identified by Moos et al. [1], in what we shall call 'The Utah Study', as being the 'best' discriminators in cDNA microarray gene expression data generated from 54 bone marrow samples obtained from childhood ALL patients presenting at the Children's Hospital at Westmead. Comparisons made by The Utah Study include childhood ALL patients having different immunophenotypic lineage (B versus T), ALL patients

* Corresponding author. Tel.: +61 2 98451205; fax: +61 2 98453078.
E-mail address: danielc@chw.edu.au (D. Catchpoole).

in different risk strata and ALL patients having the *t* (12; 21) translocation. Unlike the Utah Study, no acute myeloid leukaemia samples were included in this investigation. We describe how the differentially expressed genes identified using both analytical approaches by The Utah Study [1]; do not provide the level of distinction between these subgroups in our patient cohort. The reason for this discrepancy will be discussed. Analysis of our data with an alternative machine learning approach indicated that the expression of *endoglin*, which was identified in The Utah Study as being able to discriminate ALL lineage, was better at identifying ALL patients who eventually succumbed to the disease. This investigation highlights the difficulty in trying to identify simple biological relationships in a complex heterogeneous disease like ALL using a multidimensional gene expression data generated by microarray technology.

2. Materials and methods

2.1. Patient specimens and clinical data

Bone marrow from 54 patients with ALL presenting at the Children's Hospital at Westmead was collected between the years 1999–2003. This period corresponds to that reported by The Utah Study. All specimens, as well as the associated comprehensive clinical and follow-up data for each patient, were made available to the chief investigators for this project with the approval of and according to the guidelines established by the Children's Hospital at Westmead's, Tumour Bank Committee and institutional Human Research Ethics Committee. All new ALL patients were routinely assessed and tested for patient and disease-related prognostic indicators, specifically recording age at diagnosis, white cell count, cytogenetics with molecular assessment for translocations and DNA ploidy. All patients were subsequently treated following the Berlin Frankfurt Munster (BFM) 95 protocol.

Bone marrow aspirates in excess of diagnostic requirements were collected into EDTA tubes and subsequently snap-frozen in liquid nitrogen and stored at -80°C in the hospitals Tumour Bank. A cohort of bone marrow samples were also collected, from donors who did not have leukaemia and whose marrow has been deemed normal by light microscopy examination.

2.2. RNA isolation and microarray analysis

Total RNA was extracted from frozen bone marrow samples using Trizol LS (GibcoBRL, Integrated Science) following a modified procedure which allows the successful and reproducible extraction of intact total RNA from frozen archival bone marrow samples [2]. RNA quality was assessed by determining the $A_{260/280}$ ratio by spectrophotometry, while $3\ \mu\text{g}$ RNA was loaded onto a 1% denaturing agarose subject to electrophoresis followed by staining with ethidium bromide. Gel images were taken and the intensity

of the 28S and 18S ribosomal RNA bands determined by pixel density with the 28S:18S ratio [3]. RNA from 10 normal samples were pooled and subsequently used as control material in microarray experiments so as to remove any bias introduced by individual gene expression variation within our control population.

The techniques used for the microarray cDNA preparation, indirect fluorescence labelling and hybridization are per the techniques used at The Institute for Genome Research (TIGR, USA) [4]. We combined equal amounts of Cy5-labelled cDNA derived from ALL patient mRNA with the Cy3-labelled cDNA from pooled normal mRNA controls and mixed with $5 \times$ SSC buffer which contained 25% formamide, 0.1% SDS, $10\ \mu\text{g}$ human Cot1 DNA, $20\ \mu\text{g}$ heat-denatured ssDNA, $6\ \mu\text{g}$ polyA, and $12\ \mu\text{g}$ yeast tRNA. This cocktail was denatured for 5 min at 95°C prior to being hybridized to cDNA microarrays over night at 42°C . cDNA microarrays were supplied by the Peter MacCallum Cancer Centre and consisted of 10,500 sequence verified human genes spotted onto Telechem[®] slides. Following hybridization, the microarray was washed in a pre-warmed (50°C) $1 \times$ SSC solution containing 0.03% SDS for 5 min followed by successive 5 min washes in $0.2 \times$ SSC and $0.05 \times$ SSC at room temperature. The microarrays were scanned on an Axon II Scanner with a multi-channel image generated which was subsequently analysed with Genepix software (Axon, USA).

2.3. Data analysis

Each ALL RNA sample was hybridized to at least three separate cDNA microarray slides. Feature and background intensity was generated for each of the Cy5 (*R*) or Cy3 (*G*) channel. Ratios of local background corrected fluorescence intensity from each channel were \log_2 transformed ($\log_2 R/G$). The microarray $\log_2 R/G$ ratio data generated from Genepix was normalised using the within-print-tip-group intensity dependent location normalization (LOESS) within Bioconductor (R-package <http://www.r-project.org/>) within Bioconductor (R-package <http://www.r-project.org/>) was performed on all the microarrays. Missing values from the resultant normalized microarray data was filled using K nearest neighbour approach as per The Utah Study. Normalised data was merged for each sample replicate with mean $\log_2 R/G$ value used as the gene expression value for subsequent analyses.

The genes previously selected by The Utah Study as having discriminating gene expression signatures were identified in our microarray data set by cross matching the Genbank accession numbers following conversion to Unigene codes. Hierarchical clustering with the centred Pearson's similarity metric used by The Utah Study was performed using the acuity software package (Axon, USA). In addition, principal component analysis (PCA) was used to assess the discriminatory ability of the genes for each subgroup comparison using a multi-dimensional scaling approach. Finally, we subjected the genes to a two-tailed homoscedastic *t*-test to determine

the chance (*p*-value) of these each gene discriminating the subgroups in our ALL sample cohort.

Infoscore does not consider the actual gene expression values, but the ranking and position of values are being compared and hence, can indicate whether the data sets differ or not and how greatly they differ. The data generated using Infoscore did not demonstrate any unique attributes when compared to the *t*-test. Consequently, no data from the Infoscore analysis will be shown but will be referred to in the text only.

2.4. Artificial neural network (ANN) analysis

For this analysis we included data generated from additional samples with 64 patients being examined. The patient cohort were divided on the basis of clinical status with 56 being alive at the time of analysis and 8 who had died. In a separate comparison 63 patients were divided on the basis on cell lineage with 47 being B-lineage and 16 T-lineage. The gene expression ratios were normalized using LOESS normalization method. In order to include only high quality data in the ANN analysis, the clones which had less than 90% of features generating signal above local background across all samples were excluded. There were 4989 clones that passed this initial filter. In case of average quality greater than 0.9, the expression ratio of low quality spot in an individual sample was replaced with the average ratio of this spot across all samples. We then used PCA to reduce the dimensional-

ity of the data to the top 10 principal components as inputs for ANN. We used feed-forward resilient back-propagation multi-layer perceptron ANN with 3 layers [5]. The contribution of each gene to the classification was determined by the ANN models by measuring the sensitivity of the classification to a change in the expression level of each gene. In this way, the genes were ranked according to their significance to the classification.

3. Results

3.1. Annotation and gene selection

Both our gene dataset and that of The Utah Study were identified using Genbank accession numbers. To cross-match the equivalent genes on our microarray with The Utah Study signatures, all relevant Genbank accession codes were converted to Unigene cluster code. Our gene set was interrogated on the basis of both Unigene accession numbers and gene name. In the identified genes used to discriminate the various ALL sub groups, we found between four and seven genes were not present on our array. For the gene sets used to distinguish risk groups or Tel/AML cytogenetic status, the genes not found on our microarray consisted of expressed sequence tags and unspecified clones. With regards to the lineage comparison, gene not identified in

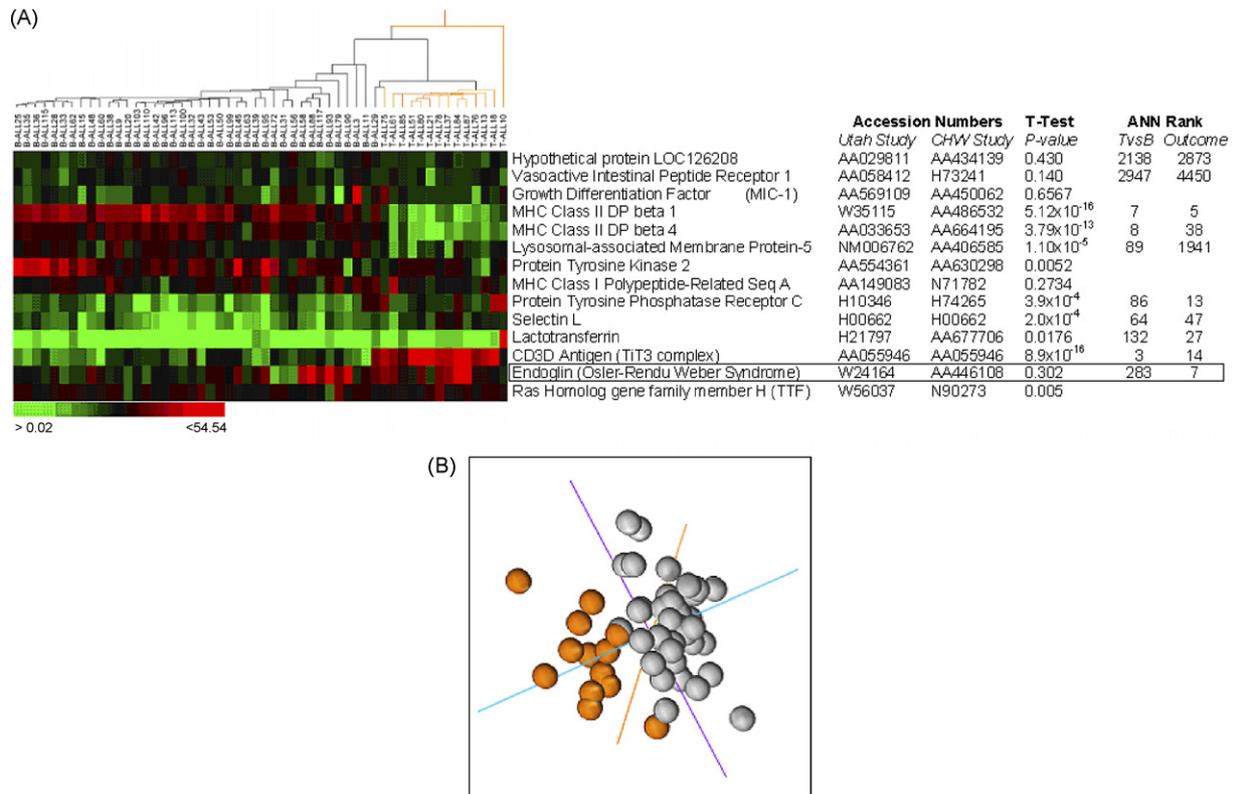


Fig. 1. Comparison of childhood ALL samples based on genes from The Utah Study scored by *t*-test reported to distinguish T & B-lineage with (A) Hierarchical cluster plot. T lineage indicated by orange dendrogram lines. Genebank annotation, *p*-value following *t*-test and the ANN ranking data for lineage and outcome comparisons are shown. (B) PCA plot. T-lineage samples are indicated with orange spheres.

our data set, and hence excluded from our validation process include *CD74* (also called MHC class II HLA DR gamma) (AA047040), *MHC class II HLA-DR-alpha* (W24610), *MHC class II HLA-DR-alpha* (W17387), *MHC class II HLA-DQ-alpha* (W67573), and *CD1e antigen* (W05301). A number of genes in the discriminating gene list were identified as ‘no significant match’ and consequently could not be validated.

3.2. Lineage subgroup: T-lineage versus B-lineage

The genes identified as best distinguishing the B-lineage and T-lineage ALL subgroups were compared in our cohort. Due to the annotation shortfalls, 14 genes identified using *t*-test and 15 genes with Infocore in The Utah Study were used to distinguish our T-ALL (14) from B-ALL (40) samples. Dendrograms generated by the hierarchical cluster analyses indicated a clear distinction between the T and B-lineage (Fig. 1A). The Utah Study indicated that the 20 best discriminating genes were identified with *p* ranging from 2×10^{-9} to 6×10^{-4} which was considered to be of ‘high significance’, although it is unclear whether a correction for multiple hypothesis testing was considered. In our cohort, only three genes reached significance when using the *t*-test and considering a Bonferroni adjustment for significance ($\alpha = 5 \times 10^{-6}$); *CD3D antigen* (AA055948), and the two *MHC class II DP* genes (AA033653, W35115). *Lactotransferrin* (H21797), *protein tyrosine phosphatase receptor type C* (H10346), *endoglin* (W24164) and *protein tyrosine kinase 2* (AA554361) indicate differential expression, but do not differentiate according to leukaemia lineage. The remaining genes did not correspond with the published paper, as no distinguishing expression pattern was observable (Fig. 1A). Principal component analysis (PCA) similarly supported the ability of The Utah Study lineage classifier to draw a distinction between T and B-ALL (Fig. 1B) with 57.0% of the variance being represented by the first three components. Gene sets identified by both the *t*-test and Infocore lead to successful segregation of samples on the basis of lineage in our samples cohort.

To further interrogate our findings, our cDNA microarray data was independently evaluated using an ANN approach. Following quality filtering, 4989 genes were assessed to identify genes best at distinguishing ALL on the basis of lineage. Genes were ranked and those used by The Utah Study were identified (Fig. 1A). Of specific interest was the rank of 283 for the gene for *endoglin*, a proliferation-associated cell membrane antigen not previously identified as being able to distinguish lymphocyte lineage. This gene was identified by The Utah Study as a potential lineage marker. However, following repeat analysis using the ANN, but comparing ALL patients on the basis of clinical outcome, *endoglin* was ranked number 7 in its ability to distinguish ALL patients who were alive at the time of analysis compared to patients who had died (Fig. 1A). Examination of the cDNA microarray gene expression ratio for *endoglin* further indicated this distinction in the 36 patients who had reached a 5-year event free survival

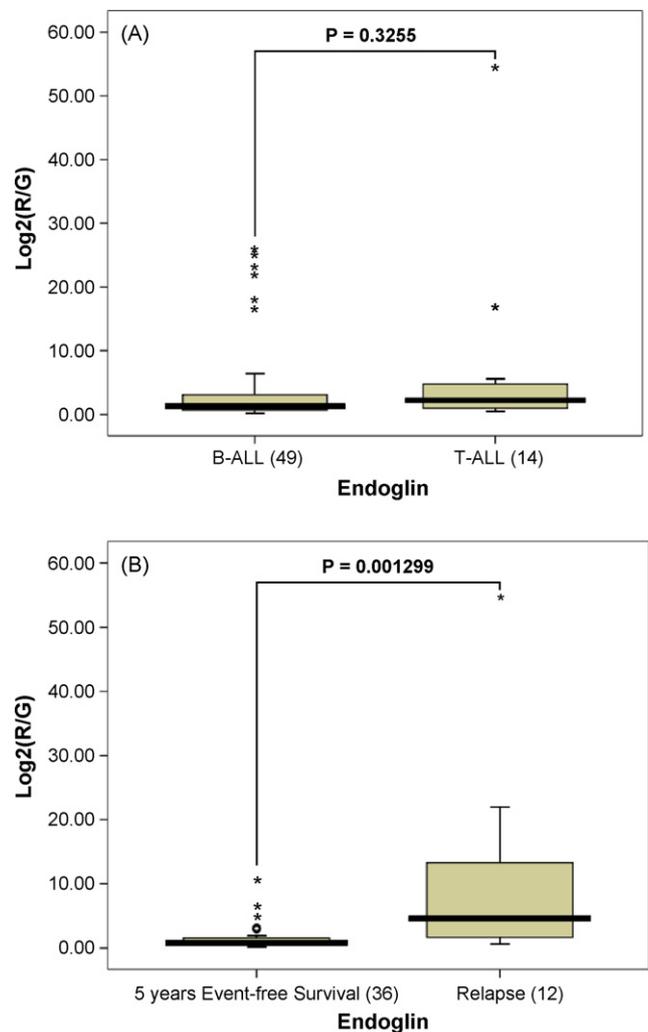


Fig. 2. Comparison of the normalised $\log_2(R/G)$ cDNA microarray values for *endoglin* in childhood ALL samples. Box plots demonstrate the distribution of values when comparing (A) cell lineage or (B) patient outcome. The bar indicates the mean value for the sample cohort, whilst the box 75% of distribution, the whiskers 25% whilst the asterisk are outliers. Of the patients who reached 5 year EFS, 29 were B-lineage with a mean $\log_2 R/G$ expression value of 1.24 and 7 were T-lineage at a mean of 3.138. For the patients who had relapsed 6 were B-lineage at a mean of 8.153 and 6 were T-lineage at 12.784.

endpoint compared to the 12 who had suffered a relapsed or died (Fig. 2A and B) whilst *t*-test confirmed a significant variations in the distribution of values between ALL patients on the basis of outcome ($p = 0.00129$). Fig. 1A indicates that the ranks of The Utah Study lineage classifier genes. Like *endoglin*, *selectin L*, *lactotransferrin* (H21797) and *protein tyrosine phosphatase receptor type C* (H10346) demonstrate a similar elevation in rank when discriminating between ALL samples on the basis of outcome. However, only *selectin L* reached significance ($p = 0.03$). By contrast, *lysosomal-associated multispinning membrane protein-5* (NM006762) did not rank highly on the basis of outcome comparisons. *Vasoactive intestinal peptide receptor 1* (AA058412), and *LOC 126208* (AA029811) were lowly ranked with both ANN

comparisons and did not appear discriminatory following this analysis.

3.3. Cytogenetic subgroup: Tel-AML positive & negative

Only five (5) Tel-AML positive B-ALL patients were present in our cohort. When compared to the Tel-AML negative (43) data slight differences were noted. Clustering of our sample cohort using the published genes from both the *t*-test (14 annotated) and Infocore (12 annotated) approaches showed no ability to distinguish B-ALL samples having the Tel-AML translocation (Fig. 3A). Expression of *glyoxylate reductase* (W39164/R83908), a gene focussed on in the published results, did not show differential expression as expected and in fact demonstrated increased expression in four of the five Tel-AML positive samples, which is the converse of the findings of The Utah Study. PCA demonstrates close relationship of the Tel-AML positive samples in both gene classifiers with 44.5% of the total variance being represented with the first three principal components in the *t*-test classifier. However, no distinction from the Tel-AML negative samples was observed (Fig. 3B). No gene was found to

distinguish Tel-AML positive and Tel-AML negative patients to a significant *p* value when tested across our samples.

3.4. Clinical subgroup: standard risk (SR) versus medium risk (MR)

Our ALL cohort were divided into classes based on risk stratification protocol used in BFM95 protocol with which these patients were treated. This is equivalent to the NCI criteria followed by The Utah Study. However, as the exact criteria was not stated, direct comparison of risk stratification subgroups was not possible. However, we used the published *t*-test and Infocore derived classifiers to distinguish the two major subgroups in our cohort, SR and MR B-ALL. Neither classifier was able to distinguish these clinical subgroups. This is illustrated by both hierarchical clustering (Fig. 3C) and PCA (Fig. 3D) where 61.7% of the total variance is represented within the first three principal components. *Protein kinase C substrate* (NM002743), the gene identified in The Utah Study, had a *p* values of 0.208 which was the second most significant of the risk group classifier. No gene was found to distinguish patients on the basis of BFM95 risk

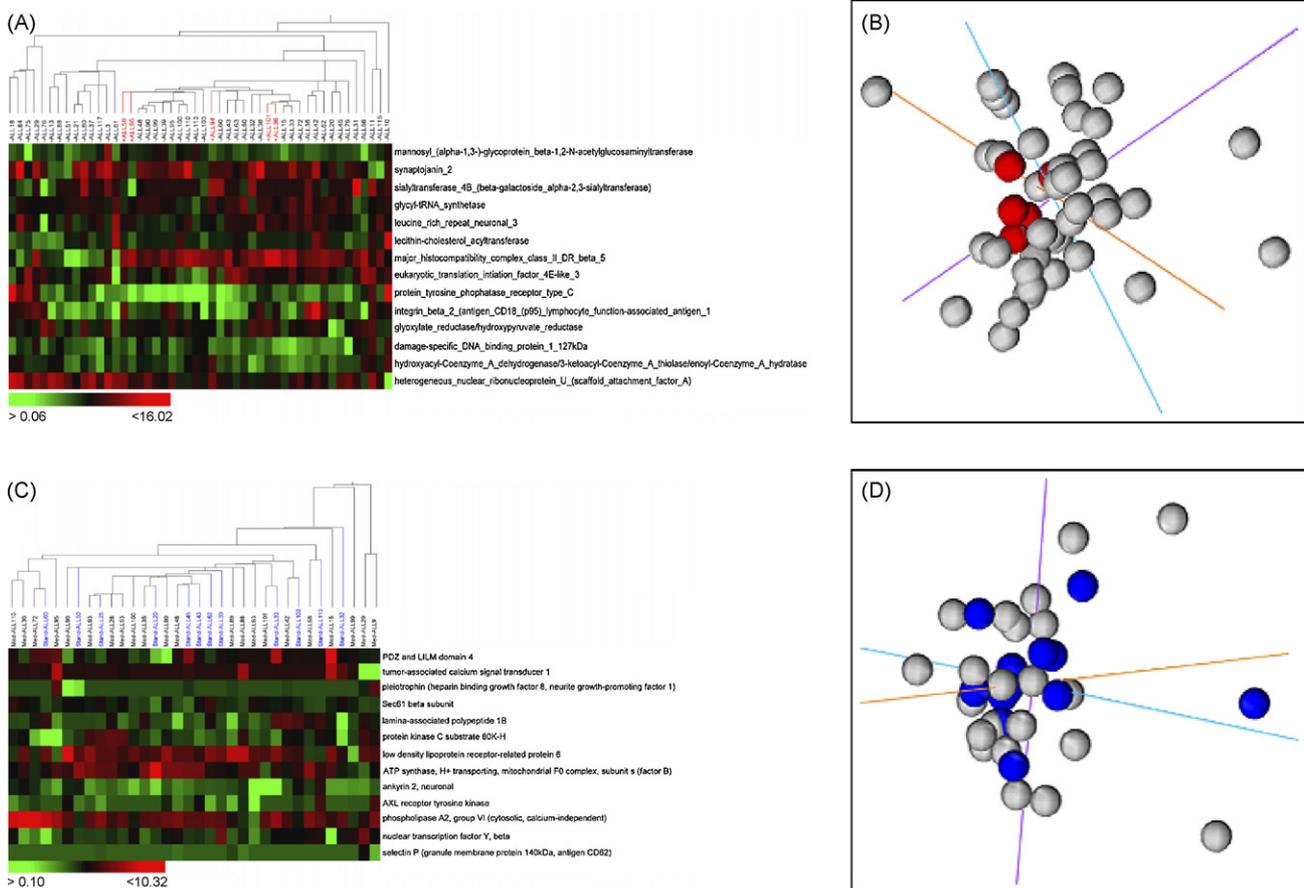


Fig. 3. Comparison of childhood ALL samples based on genes from The Utah Study scored by *t*-test reported to distinguish (A and B) Tel-AML1 cytogenetics or (C and D) risk stratification. (A and C) Hierarchical cluster plots of reported distinguishing genes which were identified by The Utah Study which were able to be cross matched in our dataset. (B and D) PCA plot. Red spheres represent samples positive for the TelAML translocation. Blue spheres represent patients stratified as standard risk (SR) according to the BFM95 protocol.

groups to a significant p value when tested across our patient cohort.

4. Discussion

Investigations such as The Utah Study aim to reduce highly dimensional data derived from complex diseases to identify small numbers of genes which distinguish patients as belonging to particular disease subgroups. If such ‘signatures’ of differentially expressed genes are to be used for diagnostic and clinical applications, such gene expression patterns need to be equally informative across sample cohorts from different sites. That is, the utility of this technology for molecular diagnosis should demonstrate a degree of resilience to variations which may be introduced during sample collection as well as platform and technical conditions under which the experiments are performed. With this in mind, we have undertaken a validation study of genes previously published as being able to discriminate different childhood ALL subtypes [1]. By making such a comparative investigation using the same cDNA glass slide microarray platform whilst using an equivalent but independently derived sample cohort from The Children’s Hospital at Westmead, Australia, we hoped to determine whether this approach to analysing microarray data will have utility in the broader diagnostic setting.

Of the 14 genes identified by The Utah Study, which were available for validation and that best discriminate B-lineage and T-lineage ALL immunophenotypes, a suitable distinction was also made within our patient cohort. A number of individual genes identified by both t -test and Infoscience approaches could, in isolation, be used to identify a B-lineage ALL from a T-lineage. *CD3* is a classic marker for T cell malignancies and is often identified as a highly significant differentially expressed gene when comparing lineages in acute lymphoblastic leukaemia. All but one identified T-lineage ALL in our cohort had highly increased *CD3* expression when compared to normal bone marrow. Similarly, the MHC histocompatibility complex gene family, whose expression was confirmed using rtPCR by the Utah Study, were overexpressed in B-lineage ALL as expected. It is most likely that the combination of these few genes is what identified the distinction evident in our cohort (Fig. 1B). None of the other genes selected by both t -test and Infoscience, as being differentially expressed in The Utah Study cohort were shown to be differentially expressed between the lineage subtypes in our dataset.

Indeed, no genes which were identified as being able to classify ALL on the basis of cytogenetic abnormality or risk stratification criteria by The Utah Study were able to distinguish the same groups in our patient cohort (Fig. 3). The low combined variance levels in the first three components of each of the PCA plots further indicates the lack of distinction between the subgroups based on the genes identified. With these comparisons, The Utah Study indicated that all of the

best discriminating genes were identified with a probability ranging from $p = 0.00001$ to 0.0002 . However, consideration of a correction for multiple hypothesis testing, such as Bonferroni, would put in doubt whether these probability values were, in reality, significant.

Despite the similarities in the sample cohort and microarray platform, differences with sample preparation and the experimental design between the two studies needs to be identified. Firstly, The Utah Study worked with freshly isolated bone marrow which was subsequently processed to isolate the mononuclear cells using Ficoll. In our case, bone marrow was obtained from the Children’s Hospital at Westmead Tumour Bank where it had been cryopreserved as whole samples by snap freezing in liquid nitrogen soon after aspiration. RNA isolation from the frozen bone marrow pellet was undertaken using a method derived specifically for this purpose [2] and yields high quality RNA [3], which, in our hands, is of better quality than that which can be achieved from cells following Ficoll isolation (data not shown). All the ALL samples within our sample cohort were identified as having high blast count, so the gene expression data should represent the leukaemic population. Nonetheless, there is the consideration that the influence of non-lymphoid and non-leukaemic nucleated white cells may create discrepancies between the two gene expression datasets.

Of greater importance to the interpretation of results from these studies is the source of the control RNA used during hybridization. The Utah Study used RNA isolated from the HL60 cell line, which is known to be sensitive to differentiation effects of culture conditions, hypoxia and confluence, and may not represent a constant background of gene expression against which to compare the ALL samples. Indeed, a number of the genes identified in this study are known to be influenced by the differentiation status of HL60 cell lines. *Endoglin* is known to be upregulated in myeloid cells undergoing macrophagic differentiation for which phorbol 12-myristate 13-acetate treated HL60 is an experimental paradigm [6], whilst *selectin-L* expression is diminished in retinoic acid treated HL60 cells [7]. Further, The Utah Study was comparing diseased cells to an attenuated diseased cell which effectively defeats their stated purpose of using microarray technology to identify relevant pathways which may provide the targets for novel therapeutics [1]. In our experiments however, bone marrow from normal (non-malignant) donors was pooled from different groups of 10 individual samples. As such, for our comparison, the ALL bone marrow tissue was compared to normal tissue representing the normal population.

The identification of differentially expressed genes within microarray data is often performed to identify key biomarkers which can be applied to diagnostic purposes, investigated for functional activity or become the target for therapeutic strategies. *Endoglin* was identified in the original microarray report as significant in the distinction between B-lineage and T-lineage ALL, and was confirmed using rtPCR on the training sample cohort. This was, however, not confirmed as differ-

entially expressed by rtPCR on an independent test set. This set only contained two T-lineage ALL samples [1]. Indeed, *endoglin* showed no differential expression which segregated with patient immunophenotype in our sample cohort either. However, examination of expression values across our sample cohort did indicate a distinct range of expression values which was unlike any of the other genes identified by The Utah Study except for the known lineage related genes CD3 and MHC gene family. It is well established that T-lineage ALL are more difficult to treat than B-lineage ALL. Hence, when making an alternative comparison involving other clinical criteria, the identification that *endoglin* expression at diagnosis could distinguish childhood ALL on the basis of patient outcome (5-year event free survival versus relapse or death) (Fig. 2A), stands to reason and reflects a unique biological role for *endoglin* in childhood ALL. This result may reflect how the microvascular environment influences patient response to therapy. Similarly, it may reflect the presence of more proliferative subpopulations of haematopoietic stem cells [8]. The prognostic potential of *endoglin* expression in bone marrow for childhood ALL should be explored further.

More significantly, the results from this study indicate that reductionist data analysis is unlikely to identify biomarkers or small subsets of genes, the expression of which will universally segregate patients into subpopulations, especially when relatively small numbers of samples are collated to represent a complex and heterogeneous diseases such as childhood ALL. For this to be achieved the high dimensionality of microarray data necessitates that samples in the numbers approximating the number of gene features on a microarray be collected before we can confidently move forward. In this case, our results with *endoglin* highlight that such reductionist approaches may oversimplify expected relationships and obscure significant biological interactions. Rather consideration of the inclusion of all clinical and biological features of childhood ALL into the data analysis methodology will potentially increase the diagnostic potential and knowledge discovery which can be gained from microarray gene expression investigations.

Acknowledgements

The Tumour Bank would like to thank Dr Andrew Holloway of the Peter MacCallum Cancer Centre for provision of the cDNA microarrays. Further, Mr Steven Wolf provided assistance through graphical preparation of figures used in this manuscript and Mrs Libby Catchpoole for proof reading. Dachuan Guo is supported by the charitable organisation, Kayaking for Kemo Kids. Daniel Catchpoole is a Research Fellow of the University of Sydney.

References

- [1] Moos PJ, Raetz EA, Carlson MA, Szabo A, Smith FE, Willman C, et al. Identification of gene expression profiles that segregate patients with childhood leukemia. *Clin Cancer Res* 2000;8:3118–30.
- [2] Barbaric D, Dalla-Pozza L, Byrne JA. A reliable method for total RNA extraction from frozen human bone marrow samples taken at diagnosis for acute leukemia. *J Clin Pathol* 2002;55(11):865–7.
- [3] Guo D, Cutri B, Catchpoole DR. The effects of RNA integrity, purity and cDNA labelling on glass slide cDNA microarray image quality. *BMC Genome Biology* 2004;5(8):13. <http://genome-biology.com/qc/2004/5/8/p13>.
- [4] Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, et al. A concise guide to cDNA microarray analysis. *Biotechniques* 2000; 29:548–62.
- [5] Wei JS, Greer BT, Steinberg S, Son C-G, Chen Q-R, Westermann F, et al. Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. *Cancer Res* 2004;64(19):6883–91.
- [6] Bellon T, Corbi A, Lastres P, Cales C, Cebrian M, Vera S, et al. Identification and expression of two forms of the human transforming growth factor-beta-binding protein endoglin with distinct cytoplasmic regions. *Eur J Immunol* 1993;23(9):2340–5.
- [7] Sjogren F, Stendahl O, Ljunghusen O. The influence of retinoic acid and retinoic acid derivatives on beta2 integrins and L-selectin expression in HL-60 cells in vitro. *Inflammation* 2000;24(1):21–32.
- [8] Chen C-Z, Li M, de Graaf D, Monti S, Gottgens B, Sanchez M-J, et al. Identification of endoglin as a functional marker that defines long-term repopulating hematopoietic stem cells. *PNAS* 2002;99(24):15468–73.