Compositional analysis gives insight into leukaemia cell lines expression profiles compared to those within patient sub-groups

Mills, K., & Blayney, J. (2018). Compositional analysis gives insight into leukaemia cell lines expression profiles compared to those within patient sub-groups. *British Journal of Haematology*, *181*(6), 847-851. https://doi.org/10.1111/bjh.14711

Published in:
British Journal of Haematology

Document Version:
Peer reviewed version

Compositional analysis gives insight into leukaemia cell lines expression profiles compared to those within patient sub-groups

Jaine K Blayney and Ken I Mills

Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, United Kingdom

Words: 975 (max 1000)
Figures and Tables: 2 (max 2)
Supplementary Tables: 1
References: 10 (max 10)

The use of cell lines has been, and continues to be, an invaluable tool for understanding the cell and molecular biology of malignant haematopoiesis. Good culture practice should involve regularly confirming authenticity to identify possible cross-contamination from other cells (Capes-Davis *et al*, 2013). However, a further issue exists: how closely does a cell line represent the parental disease process. A gene expression profile represents an integration of all of the molecular and mutational processes driven by the disease and has been used to characterise, classify, predict response and to provide prognostic information. For example, does an acute myeloid leukaemia (AML) cell line actually represent the molecular processes associated with the phenotype?

Using the power of gene expression profiling of a disease reference data we have robustly allocated a panel of commonly used leukaemia cell lines to disease sub-groups. GECA, **G**ene **E**xpression **C**ompositional **A**ssignment ((Blayney *et al*, 2016)), is an *in silico* method using compositional statistics to transfer prior knowledge from a reference data set to an unclassified query set. The approach enables the comparison of similarities in gene expression profiles across independent datasets, platforms and technologies, thus removing cross platform normalisation. GECA had been applied to a library of epithelial ovarian cell lines with respect to a reference set of solid tumours (Blayney *et al*, 2016).

In this study, the GCEA pipeline used 10,000 random gene-sets (of 500 in length) to compare the similarity of three cell line query datasets to the disease reference data set. These similarity assignments were tested for robustness, by random permutation of ranks to produce an estimated p-value, adjusted for multiple comparisons. A false discovery rate (FDR) of 0.05 was then applied.

The disease reference data set used was the MILE data set (Haferlach *et al*, 2010) (GEO: GSE13204): 2,096 samples from 17 leukaemia sub-groups. Query dataset 1 (Q1) comprised 39 haematological

cell lines as part of an Astra Zeneca project of 627 cell lines (GEO: GSE57083), profiled on the Affymetrix (Affy) HG133Plus2 platform. Query dataset 2 (Q2) was 20 profiles from 11 different cell lines, also on Affy HG133Plus2 platform; whilst Query data 3 (Q3) included six profiles representing five cell lines on the Affy HG133A platform. For Q1 and Q2, GCEA analysis was undertaken at the probe-set level (n = 54,630) with NA values removed; whilst for Q3, intersecting platform probe-set levels were used (n = 22,232) also with NA values removed. Composite aggregate assignments for each of the 18 MILE subgroups were then calculated, converted to percentages (Supplementary Table 1) and displayed as heat maps for the query data sets (Figure 1).

Several cell lines were represented in two or three of the query data sets which enabled the development of a consensus similarity across platforms; summarised in Table 1.

HL60, a commonly-used cell line showed the largest transcriptional similarity with *AML with normal karyotype + other abnormalities* (*AML-NKO*) reference group (average 46.5% similarity over Q1-Q3). Similarities were found with *AML with 11q23 /MLL* (~15%); *AML with t(8;21)* (~11.9%) and *AML with t(15;17)* (8.1%) groups. HL60 was derived from an AML FAB M2 (Collins, 1987),is responsive to ATRA differentiation therapy, similar to acute promyelocytic leukaemia (APL) and is used as a model for that disease.

K562 cells were established from a chronic myeloid leukaemia (CML) patient in blast crisis with a Philadelphia chromosome (Lozzio & Lozzio, 1975). Across Q1-3, K562 demonstrated high levels of similarity with *AML-NKO (29-64%),* with variable similarity (6.4% - 47.5%) with the *MDS* group. Surprising, and across all three query sets, no, or very little, similarity with the *CML* or *ALL with t(9;22)* group was seen.

NB4 (Duprez *et al*, 1992) and Kasumi-1 (Asou *et al*, 1991), represent *AML with t(15;17*) and *AML with t(8;21)* respectively. For NB4 cells, the dominant similarity was with the *AML-NKO* group (Q1-Q3 average of 36.7%) followed by *AML with t(15;17*) (22.4%). Kasumi-1 had the greatest similarity with *AML with t(8;21)* (37.4% in Q1, 60.6% in Q2).

U937 was most similar to *AML-NKO* (47.6%) with 24.4% similarity to *AML with t(11q23)/MLL;* U937 cells have the t(10;11)(p14;q23) translocation (Sundstrom & Nilsson, 1976).

The SKM1 and UT7 were shown to have similarity with *AML-NKO,* but in addition, GECA predicted assignments to *MDS*: SKM-1 with an average of 15% and UT-7 at 22.3%. SKM-1 cells were established from an AML patient in transformation following MDS (Nakagawa & Matozaki, 1995)

whilst the UT7 was from an AML FAB group M7 (Komatsu *et al*, 1991) but with no description of a preceding MDS phase as suggested by the GCEA results.

Interestingly REH cells, derived from an ALL at first relapse with the t(12;21) translocation, had GECA similarity of 35.6% (Q2) and 48.4% (Q3) to *ALL with t(12;21)* samples, yet only 5.5% using the Q1 REH profile. Surprisingly the Q1 REH did show 26.8% similarity to *ALL with t(1;19)*.

Overall, GCEA has identified both similarities and conflicts in the query cell line expression profiles compared to patient derived reference dataset. The majority of cell lines mapped to a core disease of *AML-NKO* probably reflecting the stage of maturation arrest occurring during myeloid differentiation. Specific and relevant expression contributions to sub-diseases were seen for *t(11q23/MLL)* U937, *t(15;17)* for NB4, *t(8;21)* for Kasumi-1 and *MDS* for SKM1. Perhaps, the most surprising result was the lack of similarity of K562, across the query data sets, for *CML* or *ALL with t(9;22)* profiles despite the confirmed presence of the Philadelphia chromosome. Some dissimilarity could be explained by different sources of cell lines or variations in laboratory protocol, whilst others could be explained by the reduced number of probesets in Q3 which used the smaller 133A array. However, this study has demonstrated that GECA can provide insight into the value, relevance and validity of using cell lines for molecular and cellular research as models of sub-types of acute leukaemia and although this should be considered when interpreting experimental data.

Acknowledgements

JB and KIM designed the study, performed the analysis and wrote the paper.

Competing interests

The authors have no competing interests

Figure and Table legends

Figure 1: Cell lines by gene expression compositional aggregate assignment with heat maps showing percentage assignment of cell lines from each of the Query datasets 1, 2 and 3 compared to the reference data set (Haferlach *et al*, 2010) by GCEA (Blayney *et al*, 2016).

Table 1: Relative influence of each disease sub-group on the molecular expression within those cell lines which were present in at least 2 of the query data-sets; each column total is equivalent to 100%.

Supplementary data table 1: The composite aggregate assignments (sheets 1, 3 and 5) and percentage of class assignments (sheets 2, 4 and 6) mapped to the disease sub-groups for each of the Query datasets (Q1-Q3).

References

Asou,H., Tashiro,S., Hamamoto,K., Otsuji,A., Kita,K., & Kamada,N. (1991) Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood*, **77**, 2031-2036.

Blayney,J.K., Davison,T., McCabe,N., Walker,S., Keating,K., Delaney,T., Greenan,C., Williams,A.R., McCluggage,W.G., Capes-Davis,A., Harkin,D.P., Gourley,C., & Kennedy,R.D. (2016) Prior knowledge transfer across transcriptional data sets and technologies using compositional statistics yields new mislabelled ovarian cell line. *Nucleic Acids Research*.

Capes-Davis,A., Reid,Y.A., Kline,M.C., Storts,D.R., Strauss,E., Dirks,W.G., Drexler,H.G., MacLeod,R.A., Sykes,G., Kohara,A., Nakamura,Y., Elmore,E., Nims,R.W., Alston-Roberts,C., Barallon,R., Los,G.V., Nardone,R.M., Price,P.J., Steuer,A., Thomson,J., Masters,J.R., & Kerrigan,L. (2013) Match criteria for human cell line authentication: where do we draw the line? *Int.J Cancer*, **132**, 2510-2519.

Collins,S.J. (1987) The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression. *Blood*, **70**, 1233-1244.

Duprez,E., Ruchaud,S., Houge,G., Martin-Thouvenin,V., Valensi,F., Kastner,P., Berger,R., & Lanotte,M. (1992) A retinoid acid 'resistant' t(15;17) acute promyelocytic leukemia cell line: isolation, morphological, immunological, and molecular features. *Leukemia*, **6**, 1281-1287.

Haferlach,T., Kohlmann,A., Wieczorek,L., Basso,G., Kronnie,G.T., Bene,M.C., de Vos,J., Hernandez,J.M., Hofmann,W.K., Mills,K.I., Gilkes,A., Chiaretti,S., Shurtleff,S.A., Kipps,T.J., Rassenti,L.Z., Yeoh,A.E., Papenhausen,P.R., Liu,W.M., Williams,P.M., & Foa,R. (2010) Clinical Utility of Microarray-Based Gene Expression Profiling in the Diagnosis and Subclassification of Leukemia: Report From the International Microarray Innovations in Leukemia Study Group. *Journal of Clinical Oncology*, **28**, 2529-2537.

Komatsu,N., Nakauchi,H., Miwa,A., Ishihara,T., Eguchi,M., Moroi,M., Okada,M., Sato,Y., Wada,H., Yawata,Y., & . (1991) Establishment and characterization of a human leukemic cell line with megakaryocytic features: dependency on granulocyte-macrophage colony-stimulating factor, interleukin 3, or erythropoietin for growth and survival. *Cancer Research*, **51**, 341-348.

Lozzio,C.B. & Lozzio,B.B. (1975) Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*, **45**, 321-334.

Nakagawa,T. & Matozaki,S. (1995) The SKM-1 leukemic cell line established from a patient with progression to myelomonocytic leukemia in myelodysplastic syndrome (MDS)-contribution to better understanding of MDS. *Leuk.Lymphoma*, **17**, 335-339.

Sundstrom,C. & Nilsson,K. (1976) Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int J Cancer*, **17**, 565-577.