Meta-analysis of Gene Expression Consistency in the A549 Cell Line

Abigail Moore

March 2019

Introduction

Human cancer-derived cell lines, such as the A549 cell line, are the most common models used for cancer treatment research (Gillet, 2013). Cell lines are used to carry out various experiments to determine the roles of genes. One significant benefit of using cell lines is the potential for a homogenous population of cells, which would support the reproducibility of research findings (Kaur, 2012). While this is an advantageous potentiality, differences in how laboratories handle samples influence gene expression results (Lorsche, 2014; Lovén, 2012). These differences make comparing results across laboratories difficult since it cannot be known what variation exists due to differences is laboratory techniques.

We sought to elucidate the significance of cell culture and RNA-sequencing techniques within the A549 cell line. This cell line is widely used for studies of both lung cancer and viruses. While there is little research that considers the impact of laboratory techniques on this cell line, one study has shown significant changes of gene expression in relation to cell culture duration (Cooper, 2016). To determine the impact of cell culture and RNA-sequencing techniques on gene expression, we applied methods established by Fasterius and Szigyarto (2018) to publicly available A549 RNA-seq data obtained via NCBI's Gene Expression Omnibus. Within their study, Fasterius and Szigyarto aimed to clarify the role of cell line heterogeneity by establishing a cell similarity score based upon single nucleotide variants, which they correlated with gene expression. We wished to expand on this analysis by determining if cell culture or RNA-sequencing techniques show greater correlation with gene expression as compared to that of similarity scores and gene expression. These findings could further shed light on the necessity of understanding and considering the impact of laboratory techniques on experimental results.

Results

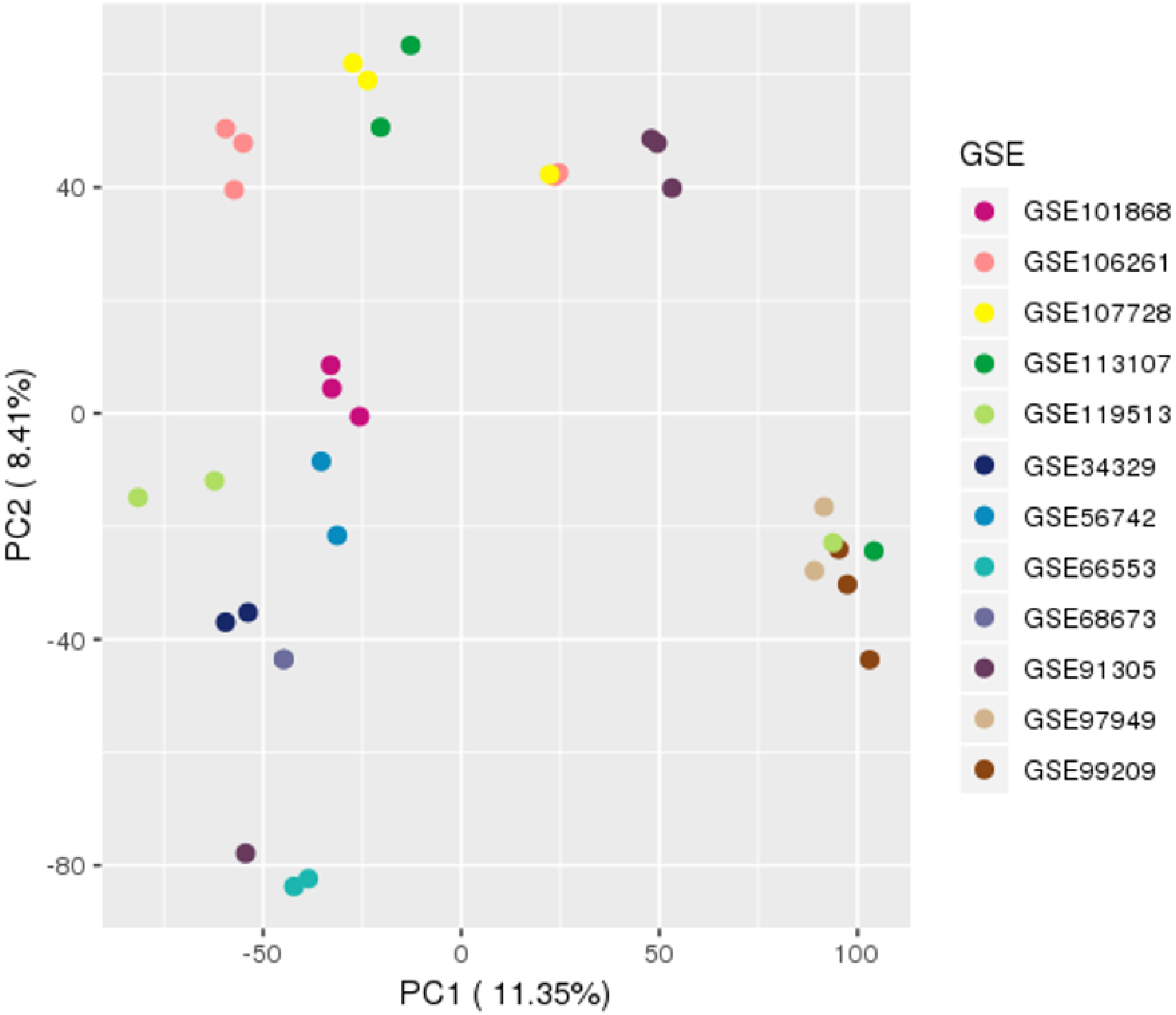


Figure 1. Principle component analysis of gene expression gathered from 12 studies (GSE). One marker represents one sample. Color of markers represents the study.

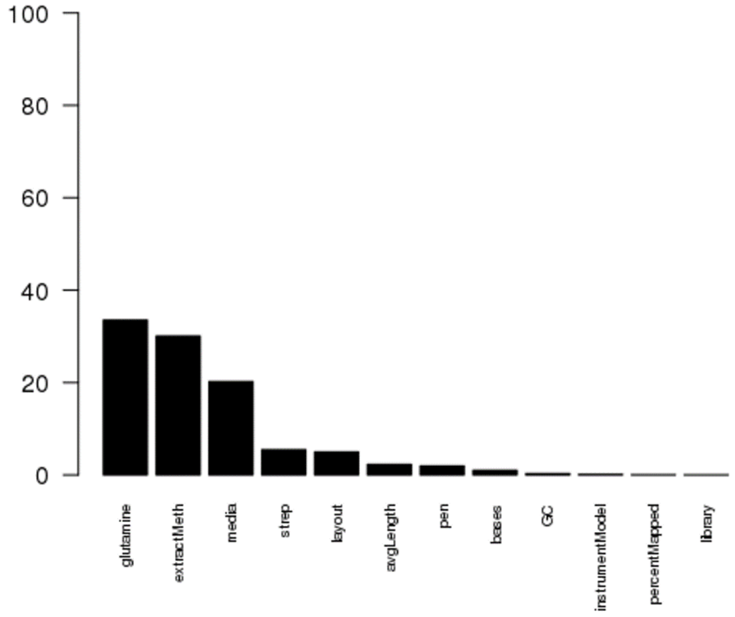
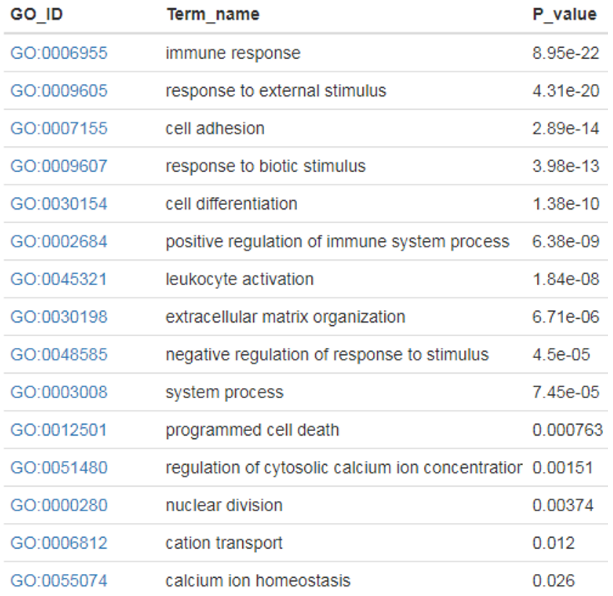


Figure 2. ANOVA of batch factors and gene expression for 12 studies. x-axis represents batch factors. y-axis represents percent contribution. extractMeth = extraction method; strep = streptomycin; avgLength = read length; pen = penicillin; GC = GC content.



GO_ID	Term_name	P_value
GO:0006955	immune response	8.95e-22
GO:0009605	response to external stimulus	4.31e-20
GO:0007155	cell adhesion	2.89e-14
GO:0009607	response to biotic stimulus	3.98e-13
GO:0030154	cell differentiation	1.38e-10
GO:0002684	positive regulation of immune system process	6.38e-09
GO:0045321	leukocyte activation	1.84e-08
GO:0030198	extracellular matrix organization	6.71e-06
GO:0048585	negative regulation of response to stimulus	4.5e-05
GO:0003008	system process	7.45e-05
GO:0012501	programmed cell death	0.000763
GO:0051480	regulation of cytosolic calcium ion concentratior	0.00151
GO:0000280	nuclear division	0.00374
GO:0006812	cation transport	0.012
GO:0055074	calcium ion homeostasis	0.026

Figure 3. Gene ontology terms identified across 12 studies.

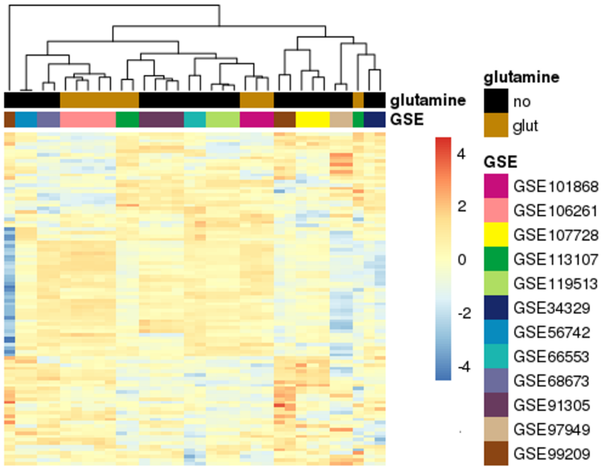


Figure 4. Differential gene expression across 12 studies with respect to study (GSE) and the batch factor with the highest percent contribution (absence [no] and presence [glut] of glutamine).

Methods

Data collection

A549 RNA-seq and meta data were acquired from NCBI GEO5 via NCBI E-utilities and R packages GEOquery and SRAdb. Cell culture meta-data required further manual curation. Query results were filtered to include untreated, wild type A549 cells, which yielded 34 samples from 12 studies. Differentially expressed genes were analyzed for lung cancer prognostic markers obtained from the Human Protein Atlas and over-representation of gene ontology terms with g:Profiler.

Differential expression analysis

Salmon9, tximport10 and edgeR11 were used to perform gene expression estimation, normalization and differential expression analysis with thresholds of false discovery rate (FDR) ¡ 0.05 and fold change (FC) $>$ 2.

References

Cooper, J. R., Abdullatif, M. B., Burnett, E. C., Kempsell, K. E., Conforti, F., Tolley, H., . . . Davies, D. E. (2016). Long Term Culture of the A549 Cancer Cell Line Promotes Multilamellar Body Formation and Differentiation towards an Alveolar Type II Pneumocyte Phenotype. PLoS ONE, 11(10), e0164438. http://doi.org/10.1371/journal.pone.0164438

Fasterius, E., Szigyarto, C. A. (2018). Analysis of public RNA-sequencing data reveals biological consequences of genetic heterogeneity in cell line populations. Scientific Reports, 8(1). doi:10.1038/s41598-018-29506-3

Gillet, J.-P., Varma, S., Gottesman, M. M. (2013). The Clinical Relevance of Cancer Cell Lines. JNCI Journal of the National Cancer Institute, 105(7), 452–458. http://doi.org/10.1093/jnci/djt007

Kaur, G., Dufour, J. M. (2012). Cell lines: Valuable tools or useless artifacts. Spermatogenesis, 2(1), 1–5. http://doi.org/10.4161/spmg.19885

Lorsch, J. R., Collins, F. S., Lippincott-Schwartz, J. (2014). Fixing problems with cell lines: Technologies and policies can improve authentication. Science (New York, N.y.), 346(6216), 1452–1453.
http://doi.org/10.1126/science.1259110

Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., . . . Young, R. A. (2012). Revisiting Global Gene Expression Analysis. Cell, 151(3), 476–482. http://doi.org/10.1016/j.cell.2012.10.012