Signatures of non–small-cell lung cancer relapse patients: differential expression and gene network analysis

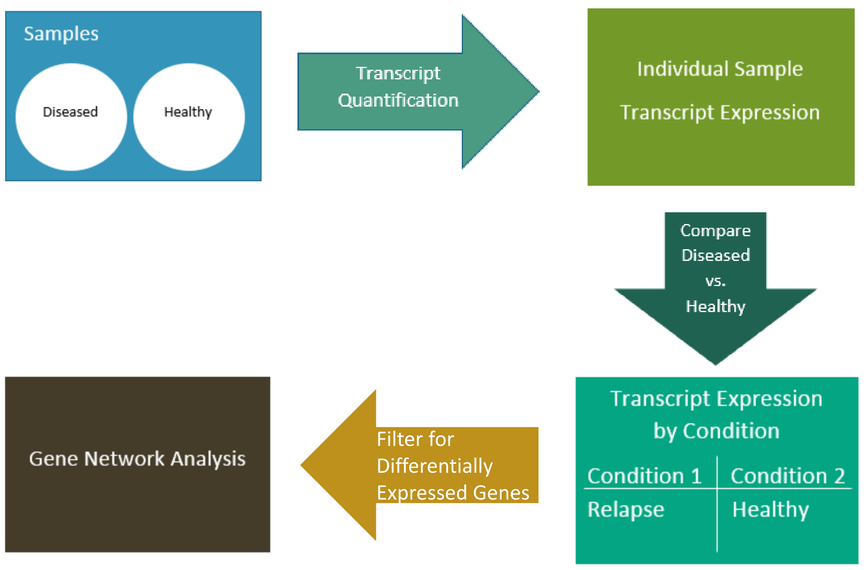
Abigail E. Moore[1], Brandon Zheng[2], Patricia M. Watson[3], Robert C. Wilson[3], Dennis K. Watson[3], Paul E. Anderson[4]
[1]Hampshire College, [2]Bard College, [3]Medical University of South Carolina, [4]College of Charleston

Introduction

Lung cancer is both the second most represented cancer diagnosis and the leading cause of cancer death within the United States[1]. Despite the high occurrence of non–small-cell lung cancer (NSCLC), 30 to 55% of patients relapse after curative resection, and the 5-year relative survival rate is 15 to 21%[1]. Biomarker programs help select patients who may benefit from a given drug, thereby high costs of cancer medication and cancer drug failures may be reduced[2,3]. We identified differentially expressed genes that are significantly linked to disease outcomes, which could lead to investigations of how these candidate biomarkers function to impact risk of relapse.

Methods

- NSCLC RNA samples were taken from 38 patients,
  - clinical outcomes were determined by the American College of Surgery Oncology Group [4];
  - 20 were diagnosed as disease-free, and 18 as relapse patients.
- RSEM[5] was used to assemble and quantify the abundance of transcripts, which produced gene–count-based expression.
- EBSeq[6] was used to perform median normalization and differential expression analysis with a threshold of false discovery rate (FDR) < 0.05.
- Differentially expressed genes were analyzed for over-representation of protein complexes, gene ontology terms and pathways with ConcensusPathDB[7].



Results

Empirical Bayesian methods[6] identified 122 differentially expressed genes (FDR < 0.05). Among these genes, only one protein complex-based set was over-represented, G protein-coupled receptor ligand. This included 7 of our differentially expressed genes, 4 of which have suggested implications in NSCLC prognosis: Existing research indicates that overexpression of TGF-α, IGF1, or IGF2 predicts poor survival, which is consistent with the gene expression of our relapse group[8-10]. Interestingly, while overexpression of HGF is also associated with poor survival, we saw higher expression of HGF in the disease-free patients[11-13].

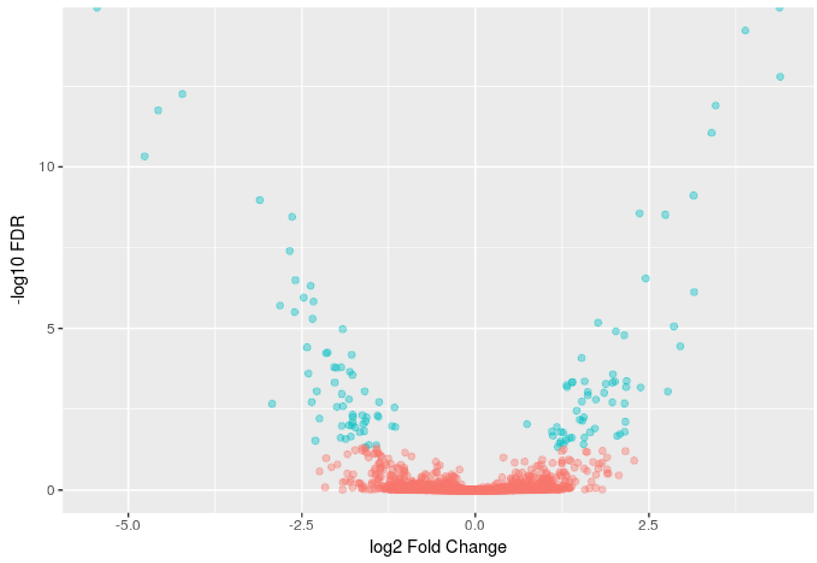


Figure 1 | Volcano plot of differentially expressed genes (blue) and similarly expressed genes (red).

Gene	PPDE	Real FC	R Mean	DF Mean
CPS1	1	21.0382203	3861.452147	183.53510
UPK1B	1	20.8719792	1640.955927	78.61052
KLK11	1	14.8335518	1476.094288	99.50118
CD70	1	11.0330937	161.379705	14.61778
PI3	1	10.5927592	2248.496323	212.25824
ADGRB1	1	8.8462140	885.679398	100.11073
KIAA0087	1	6.6737762	118.581664	17.75980
COMP	1	5.1584334	2448.075614	474.56928
SLC7A14	1	0.1605313	29.495441	183.78865
C11orf70	1	0.1568151	55.741074	355.51117
WNT11	1	0.1163370	63.640695	547.11328
SHISA3	1	0.0536560	55.093038	1026.95836
PRR4	1	0.0421482	48.838449	1158.95959
SEC14L5	1	0.0367918	9.453563	257.20964
LGALS4	1	0.0228392	28.253722	1237.50056

Table 1 | Top 15 differentially expressed genes (FDR < 0.05). Posterior probability of differential expression (PPDE); fold change (FC); relapse and disease-free normalized expression (R & DF means).

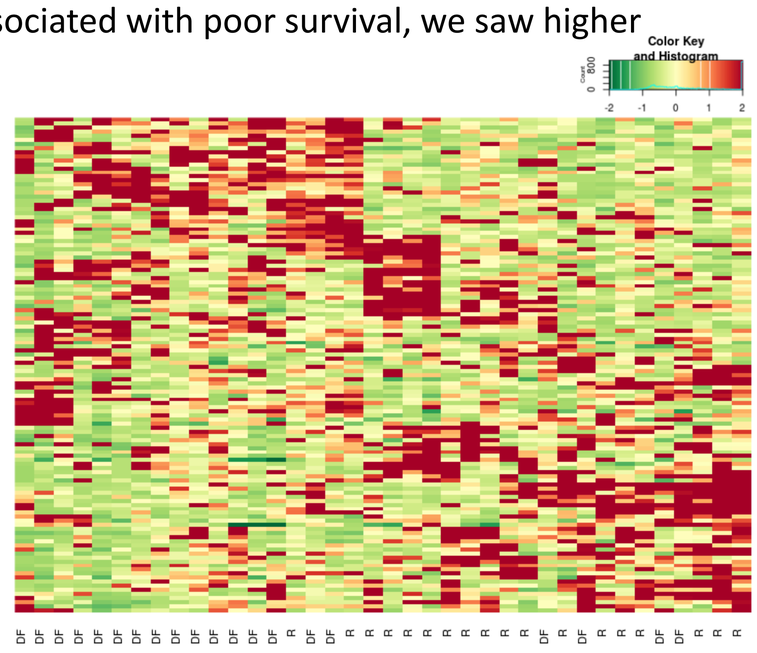


Figure 2 | Heatmap of normalized gene–count-based expression. Disease-free (DF); relapse (R). 122 differentially expressed genes on y-axis.

A

GO term	Category, level	P-value	Q-value
GO:0008284 positive regulation of cell proliferation	BP 5	4.66e-06	0.00142
GO:0019372 lipoxygenase pathway	BP 5	6.95e-05	0.0106
GO:0099503 secretory vesicle	CC 5	0.000236	0.00189
GO:0033293 monocarboxylic acid binding	MF 5	0.000267	0.00479
GO:0004252 serine-type endopeptidase activity	MF 5	0.000274	0.00479

B

Pathway	Source	P-value	Q-value
GPCR signaling-G alpha i	INOH	6.77e-05	0.00636
GPCR signaling-pertussis toxin	INOH	6.77e-05	0.00636
Prostaglandin Synthesis and Regulation	Wikipathways	8.8e-05	0.00636
GPCR signaling-cholera toxin	INOH	9.73e-05	0.00636
GPCR signaling-G alpha s Epac and ERK	INOH	0.000113	0.00636

Table 2 | **(A)** Top five level five gene ontology (GO) terms. **(B)** Top five pathways. Biological process (BP); cellular component (CC); molecular function (MF); Integrating Network Objects with Hierarchies database (INOH).

Conclusions

Identifying NSCLC patients at risk of recurrence is crucial in cancer research. Our analyses identified 122 differentially expressed genes among disease-free and relapse NSCLC patients, including known lung cancer-related genes and new candidate biomarker genes that are involved in the diverse processes related to NSCLC development. Future research in alternative splicing and the development of a predictive model based on our results could support methods of identifying individual recurrence risk.

References

1 American Cancer Society. (2017) Cancer Facts & Figures 2017. Atlanta: American Cancer Society
2 Goossens, N., Nakagawa, S., Sun, X., & Hoshida, Y. (2015). Cancer biomarker discovery and validation. *Translational Cancer Research*, *4*(3), 256–269.
3 Workman, P., Draetta, G. Schellens, J., Bernards, R. (2017). How Much Longer Will We Put Up With $100,000 Cancer Drugs? *Cell*. 168(4):579-583.
4 Anderson, P. *et al*. (2014). Predictive modeling of lung cancer recurrence using alternative splicing events versus differential expression data. *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, 1–8.
5 Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323.
6 Leng, N., *et al*. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, 29(8), 1035–1043.
7 Herwig, Ralf, *et al*. "Analyzing and interpreting genome data at the network level with ConsensusPathDB." (2016). *Nature Protocols*, vol. 11, no. 10, pp. 1889–1907.
8 Wu, H., Liu, Y., Shu, X. O., & Cai, Q. (2016). MiR-374a suppresses lung adenocarcinoma cell proliferation and invasion by targeting TGFA gene expression. *Carcinogenesis*, 37(6), 567–575.
9 Kim, J.-S., Kim, E. S., Liu, D., Lee, J. J., Solis, L., Behrens, C., … Lee, H.-Y. (2014). Prognostic Implications of Tumoral Expression of Insulin-like Growth Factor (IGF)-1 and -2 in Patients with Non-small Cell Lung Cancer. *Clinical Lung Cancer*, 15(3), 213–221.
10 Velcheti, V., Ramaswamy, G. Insulin-Like Growth Factor and Lung Cancer. (2006) *Journal of Thoracic Oncology,* 1(7). Pp. 607 – 610.
11 Ichimura, E., Maeshima, A., Nakajima, T., & Nakamura, T. (1996). Expression ofc-met/HGF Receptor in Human Non-small Cell Lung Carcinomasin vitroandin vivoand Its Prognostic Significance. *Japanese Journal of Cancer Research*, 87(10), 1063-1069.
12 Olivero, M., Rizzo, M., Madeddu, R., Casadio, C., Pennacchietti, S., Nicotra, M. R., … Di Renzo, M. F. (1996). Overexpression and activation of hepatocyte growth factor/scatter factor in human non-small-cell lung carcinomas. *British Journal of Cancer*, 74(12), 1862–1868.
13 Siegfried, J., Weissfeld, L., Singh-Kaw, P., Weyant, R., Testa, J., Landreneau, R. (1997). Association of Immunoreactive Hepatocyte Growth Factor with Poor Survival in Resectable Non-Small Cell Lung Cancer. *Cancer Res* 57(3), 433-439

Acknowledgments

Thank you to the College of Charleston Omics REU—supervised by the National Science Foundation DBI Award 1359301—the Medical University of South Carolina (MUSC) for providing the dataset, and Hampshire College's Dr. Lucy, Denice O'Neill, and Justine Salton funds.