

Metagenes Associated with Survival in Non-Small Cell Lung Cancer

Egon Urgard¹, Tõnu Vooder¹, Urmo Võsa¹, Kristjan Välk¹, Mingming Liu², Cheng Luo¹, Fabian Hoti³, Retlav Roosipuu⁴, Tarmo Annilo¹, Jukka Laine⁵, Christopher M. Frenz⁶, Liqing Zhang² and Andres Metspalu^{1,7}

¹Department of Biotechnology, University of Tartu, 23 Riia, 51010 Tartu, Estonia. ²Department of Computer Science, Virginia Tech., 2202 Kraft Drive, Blacksburg, VA 24060, USA. ³Department of Information and Computer Science, Aalto University School of Science, Otakaari 1, FI-02150 Espoo, Finland. ⁴Department of Pathology, Tartu University Hospital, 1a L. Puusepa, 50406 Tartu, Estonia. ⁵Department of Pathology, University of Turku and Turku University Hospital, FIN-20520 Turku, Finland. ⁶Department of Computer Engineering Technology, New York City College of Tech., 300 Jay St. Brooklyn, NY 11201, USA. ⁷Estonian Genome Center, University of Tartu, 61b Tiigi, 50410 Tartu, Estonia. Corresponding author email: egonur@gmail.com

Abstract: NSCLC (non-small cell lung cancer) comprises about 80% of all lung cancer cases worldwide. Surgery is most effective treatment for patients with early-stage disease. However, 30%–55% of these patients develop recurrence within 5 years. Therefore, markers that can be used to accurately classify early-stage NSCLC patients into different prognostic groups may be helpful in selecting patients who should receive specific therapies.

A previously published dataset was used to evaluate gene expression profiles of different NSCLC subtypes. A moderated two-sample t-test was used to identify differentially expressed genes between all tumor samples and cancer-free control tissue, between SCC samples and AC/BC samples and between stage I tumor samples and all other tumor samples. Gene expression microarray measurements were validated using qRT-PCR.

Bayesian regression analysis and Kaplan-Meier survival analysis were performed to determine metagenes associated with survival. We identified 599 genes which were down-regulated and 402 genes which were up-regulated in NSCLC compared to the normal lung tissue and 112 genes which were up-regulated and 101 genes which were down-regulated in AC/BC compared to the SCC. Further, for stage Ib patients the metagenes potentially associated with survival were identified.

Genes that expressed differently between normal lung tissue and cancer showed enrichment in gene ontology terms which were associated with mitosis and proliferation. Bayesian regression and Kaplan-Meier analysis showed that gene-expression patterns and metagene profiles can be applied to predict the probability of different survival outcomes in NSCLC patients.

Keywords: non-small cell lung cancer, microarray, gene expression pattern, Kaplan-Meier curve, TNM stage, metagenes

Cancer Informatics 2011:10 175–183

doi: [10.4137/CIN.S7135](https://doi.org/10.4137/CIN.S7135)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Lung cancer is one of the most common cancers worldwide and the leading contributor to cancer deaths, having one of the lowest survival rates within 5 years after diagnosis.^{1,2} Lung carcinomas are usually classified as small-cell lung carcinomas (SCLC) and non-small cell lung carcinomas (NSCLC). NSCLC accounts for 80% of all lung cancer cases. The most common histological types of NSCLC are squamous cell carcinoma (SCC), adenocarcinoma (AC) and its subtype bronchioloalveolar carcinoma (BAC).

Tumor stage according to TNM classification remains the strongest predictor of survival in lung cancer until now. The TNM staging system is based on tumor size, involvement of lymph nodes (nodal status) and presence or absence of metastases.^{3,4} However, it is not based upon intrinsic biological differences between tumor cells and does not provide a sufficiently accurate prognosis.⁵ 25% to 30% of patients are diagnosed with early-stage (ie, stage I and II) disease and are treated primarily by surgical resection. However, 30%–55% of these patients develop recurrence within 5 years, indicating the existence of biological variation and heterogeneity among patients' tumors.⁶ Therefore, markers that can be used to accurately classify early-stage NSCLC patients into different prognostic groups may be helpful in selecting patients who should receive specific therapies.

Recently, gene expression profiling has been used to identify patients with high risk of relapse.^{7–10}

We investigated gene expression signature of NSCLC by microarray analysis and report a gene expression pattern associated with subtype. We also defined for stage Ib patients the survival rates for a 1000 day cut-off together with metagenes potentially associated with a survival.

Materials and Methods

Patients and tumor samples

The Ethical Committee of Tartu University approved the sample collection based on signed informed consent with the patients. Normal samples were collected from tissues adjacent to the patients' tumors, and were confirmed to be non-cancerous by pathologists. The histological classification of the carcinomas was conducted according to the standards of the World Health Organization (WHO) classification method for carcinoma.^{11–14} Eighty five patients underwent

surgical resection and the tumors were pathologically confirmed as NSCLC pulmonary carcinoma in Tartu University Hospital between November 2002 and December 2006. Of the 85 tumors, 62 were SCCs and 23 were BCs and ACs. According to the guidelines of the American Joint Committee on Cancer¹⁵ the patients were staged after the surgery (Supplemental Table 1).

Statistical analysis of gene expression data

Quantile-normalized and log-transformed expression data was obtained from our previous study,¹⁶ which used Illumina Sentrix BeadChip (HumanWG-6_V2) to profile gene expression. Aforementioned array provides genome-wide transcriptional coverage of well-characterized genes, gene candidates and splice variants, with a significant portion targeting well-established sequences of 18,072 genes. Our initial dataset consisted of 85 lung tumor samples and 21 adjacent cancer-free lung samples (Supplemental Table 1).

Patients who had received preoperative chemotherapy were excluded from differential gene expression analyses leaving 78 tumor samples and 20 adjacent control samples.

Moderated two-sample t-test from R package LIMMA¹⁷ was used to find differentially expressed genes between sample groups, using FDR-corrected *P*-value cut-off 0.05 and fold-change 2. The approach was used to analyze differential gene expression between all tumor samples (*n* = 78) and adjacent controls (*n* = 20), between SCC samples (*n* = 58) and AC/BC samples (*n* = 20), as well as between stage I tumor samples (*n* = 56) and all the other tumor samples (*n* = 22).

The differentially expressed genes and all samples were clustered hierarchically using Pearson correlation distance with average linkage and visualized using a heatmap. Pathway enrichment analyses were carried out using GeneCodis 2.0.^{18,19}

Bayesian analysis to assess risk genes in stage Ib patients

We restricted our analysis to stage Ib patients only, since preliminary analysis suggested that gene-expression patterns associated with survival differ in different stages and stage Ib group was the largest group



(48 patients altogether, but clinical data available for only 46 of them). From the survival times we identified two distinct patient groups, a group with short survival (<1000 days, $n = 20$) and a group with long survival (>1000 days, $n = 26$) (Fig. 2A). To reduce the number of genes, 500 gene-clusters or “metagenes”, were formed from the 5000 genes with highest variance among the 46 patients. Clustering of the genes is performed using complete-linkage hierarchical clustering,^{20,21} where the distance between two genes is defined as the correlation coefficient between gene-expression values. The constructed metagenes are summarized by the mean of the genes in the cluster. Thus genes demonstrating similar variation between the subjects are summarized as a single “metagene”.

The association between the metagenes and the two groups was analyzed with a sparse Bayesian probit model for binary response variables.^{22–24} According to the model, the probability of being in the short survival group is given by $\Phi\left(\sum_{j=1}^{500} B_j x_{i,j}\right)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, B_j is the coefficient of the linear effect of metagene j , and $x_{i,j}$ is the value of metagene j for patient i for $i = 1, \dots, 46$. The model parameters (B_j) were estimated using a Markov chain Monte-Carlo (MCMC) approach together with sparsity inducing priors. For the details of the choice of prior distributions and sampling, see Hoti and Sillanpää 2006.²³ For each of the 500 metagenes, we calculated the posterior probability that the metagene is included in the model, where being included in the model is defined as having a normalized effect greater than 0.1. The normalized effect of metagene j is defined as B_j divided by the standard deviation of the values of metagene j . Thus, for metagenes with a higher variance, a smaller effect is allowed. Using the same inclusion criteria, we calculated the posterior distribution of the number of genes included in the model. To take into account the possibility that the chain converges to a local solution, we combined 20 independent MCMC runs into a single analysis. Each model was run for 30,000 rounds, and the first 20,000 rounds were removed as burn-in rounds. From the remaining 10,000 rounds, every 10th round was saved and used in the analysis. The ability of the model to predict the survival outcome of new patients was evaluated by leave-one-out

cross-validation, where one patient at a time is used to validate the model (test person), while the remaining 45 patients are used to build up the model. For each of the 46 models, the gene selection and metagene construction was performed based on the 45 patients only, and the test person was assigned to the survival group with the higher posterior probability (corresponds to cut off value 0.5).

The performance of the method was further evaluated by the ROC curve, constructed by plotting the false positive rate (sensitivity) against the true positive rate (1-specificity) for different cut off values in the range (0.1). In addition to drawing the empirical curve, we report the AUC (area under the ROC curve) value together with the P -value corresponding to the event $AUC > 0.5$. Kaplan-Meier survival curves were drawn to visually compare the survival rates in the model-predicted high- and low-risk groups. The significance of the difference was evaluated with the log-rank test.

Quantitative RT-PCR

To validate the gene expression levels detected with microarray analysis, qRT-PCR was performed for the four up-regulated genes: *TTK* (Dual specificity protein kinase), *CCNB2* (Cyclin B2), *BUB1B* (budding uninhibited by benzimidazoles 1 homolog beta (yeast)), *PTTG2* (pituitary tumor-transforming 2), and four down-regulated genes: *CLIC5* (Chloride intracellular channel protein 5), *GPR116* (G-protein coupled receptor 116), *AGER* (*RAGE*) (advanced glycosylation end product-specific receptor), *TNNC1* (troponin C type 1). We chose *ESD* (esterase D/formylglutathione hydrolase) as the endogenous reference for qRT-PCR because it has been previously identified as invariant in clinical lung cancer specimens.^{25,26} The transcripts were amplified using Maxima SYBR Green /ROX qPCR Master mix (Fermentas) and ABI Prism 7900HT (Applied Biosystems). Eight sample pairs (tumor sample with corresponding normal lung) that were present in array and three sample pairs which were not (independent validation set), were used in the qRT-PCR experiment. Relative gene expression levels were calculated using the relative quantification method (Applied Biosystems). For detailed information see supplemental method and supplemental Table 2.



Results and Discussion

Identification of two gene expression patterns and correlation with NSCLC subtypes

We identified 599 genes which were down-regulated and 402 genes which were up-regulated in NSCLC compared to the normal lung tissue (Supplemental Table 3 and supplemental Fig. 1). According to Genecodis 2.0 analyses the main up-regulated processes in cancer were not only related to mitosis, cell division, DNA replication, blood vessel development, keratinocyte differentiation and epidermis development (Supplemental Table 4). The number of various down-regulated processes is much larger including immune response, signal transduction, cell to cell adhesion, cell surface reception linked signaling pathway, cell differentiation and others (Supplemental Table 5). To find differentially expressed genes between lung cancer subtypes, we used only tumor samples of different subtypes as sample groups. Because of the small sample size and similar histology²⁷ of AC and BC samples we used these in analyses as one group. We identified 112 genes which were up-regulated and 101 genes which were down-regulated in AC/BC compared to the SCC (Supplemental Fig. 2). Some of the genes showing largest fold changes in our dataset are belonging to keratin gene family, which overexpression is shown to be SCC specific.²⁸

We also carried out analyses to identify genes which may distinguish NSCLC stages. Because most of our sample group consisted of early stage (Ia and Ib) tumor samples and there were limited number of later stage samples, we compared only I stage tumor samples with all the others. We did not find any genes which were significantly overexpressed in one of the sample groups after multiple testing correction. This may suggest that although TNM staging system reflects the clinical status of tumor, it may not be the best tool to assess the underlying biological properties of cancer caused by aberrant gene expression.

The expression of the four up-regulated and four down-regulated genes (see supplemental data Table 2) were confirmed using eight pairs of normal and tumor tissue samples from microarray analysis sample set and further validated using three sample pairs that were not analyzed using microarray previously (Fig. 1A).

The results were quantified using $2^{-\Delta\Delta Ct}$ method and *ESD* as an endogenous reference. Significant correlation was observed between log₂-transformed array signal intensities and qRT-PCR Ct values, showing the consistency between qRT-PCR and microarray data (Fig. 1B).

Among the up-regulated genes we validated, *CCNB2* and *TTK* are involved in M phase of the cell cycle,^{29–31} which indicate cell proliferation is involved in carcinogenesis. *AGER (RAGE)*, another gene we validated, showed a statistically significant down-regulation in the gene expression values between NSCLC and normal lung tissues, which is consistent with the earlier findings that down-regulation of *AGER (RAGE)* supports NSCLC.³²

Identifying genes associated with survival in stage Ib patients

To identify prognostic markers for high-risk patients with early-stage disease, the Bayesian model was first applied to both Ia and Ib patients (results not shown). However, the model-based prediction into low and high survival risk groups was clearly less accurate for Ia patients than for Ib patients. This suggests that gene-expression profiles associated with survival are different between stages Ia and Ib. A possible explanation for why the model worked for stage Ib but not for stage Ia patients could be that the number of patients in stage Ib (n = 46) was larger than that in stage Ia (n = 14).

Since the largest group of samples in our dataset were from patients with tumor stage Ib, we next focused to this subset of patients. A sparse Bayesian regression model was constructed using all 46 stage Ib patients. The number of influential metagenes supported by the model was 4 (mode) with a 95% credible interval of 1–10. The top two metagenes (*MAGEA9B* (melanoma antigen family A, 9B), *MAGEA10* (melanoma antigen family A, 10), and *NLRP2* (NLR family, pyrin domain containing 2)) and (*CYP3A5* (cytochrome P450, family 3, subfamily A, polypeptide 5), *AKR1B1* (aldo-keto reductase family 1, member B1), *FAM46B* (family with sequence similarity 46, member B)) (Fig. 2A) were included into the model with 11% and 10% posterior probabilities, respectively. For the remaining metagenes, the posterior probabilities were less than 5%. According to our model, genes

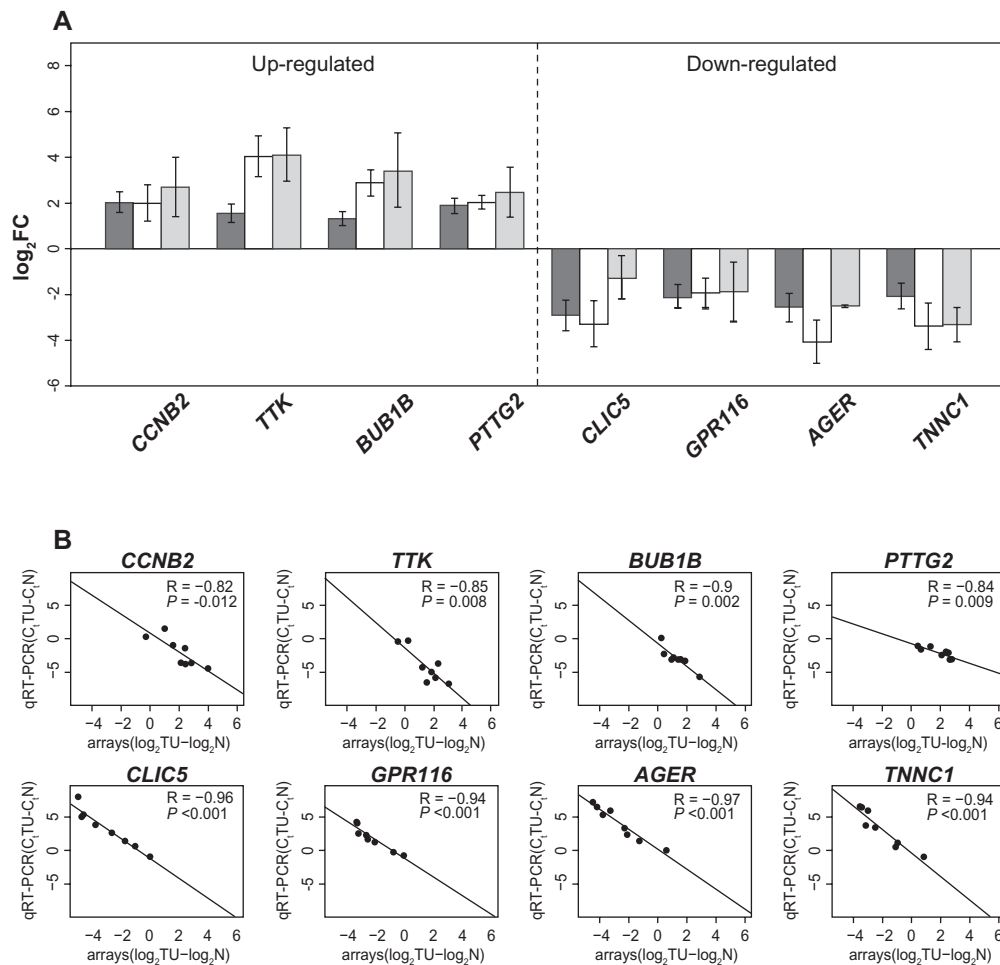


Figure 1. **A)** Validation of microarray data with qRT-PCR of four up-regulated and down-regulated genes. ■, average log fold-change for paired lung cancer samples on microarray (n = 21), □, qRT-PCR average log fold-change for the lung cancer sample pairs that were presented on microarray (n = 8), ▨, qRT-PCR average log fold-change for the lung cancer sample pairs that were not presented on the microarray (n = 3). Error bars indicate the standard error of the mean (SEM). **B)** Correlation between array $\log_2(\text{signal}_{\text{tumor}}) - \log_2(\text{signal}_{\text{normal}})$ and qRT-PCR $\Delta\Delta\text{Ct}$ for validated genes using same sample pairs as previous graph (n = 8). Pearson correlation coefficients (R), correlation test p-values, and best-fitting (least squares) lines are shown.

within these metagenes are associated with a survival prognosis of stage Ib patients, and are therefore potential targets for further research. In the SCC group, high gene expression values of the first top metagene are associated with long survival (>1000 days), while SCC group with low values are associated with short survival (<1000 days) (Fig. 2B).

Metagenes *MAGEA9B* and *MAGEA10* have been involved in congenital dyskeratosis,³³ and *NLRP2* may be involved in another type of familial imprinting disorder.³⁴ *CYP3A5*, located in the second top metagene, contains one of the most commonly detected polymorphisms that influences the efficacy of vinorelbine-based therapies to treat NSCLC.³⁵ However, the exact role of these metagenes in tumorigenesis requires further studies.

Identifying stage specific survival-associated genes would require fitting separate models or adopting a model that allows state-specific gene-effects. For example, the binary classification-tree approach³⁶ allows one to combine very different gene-expression profiles (leaves of the tree) into the same class. We used the cut-off value of 1000 days, since at that point there exists a clear gap (757 days to 1225 days) in the survival times of Ib patients. As shown in Fig. 2A, there are 3 patients who died after the cut-off that may have a negative effect on the quality of the data, ie, if the censored times may turn out to be much longer. In Potti et al 2006, the analysis was based on a cohort consisting of NSCLC patients with a recurrence within 2.5 years and patients with no recurrence within 5 years. Also, their analysis differs from ours

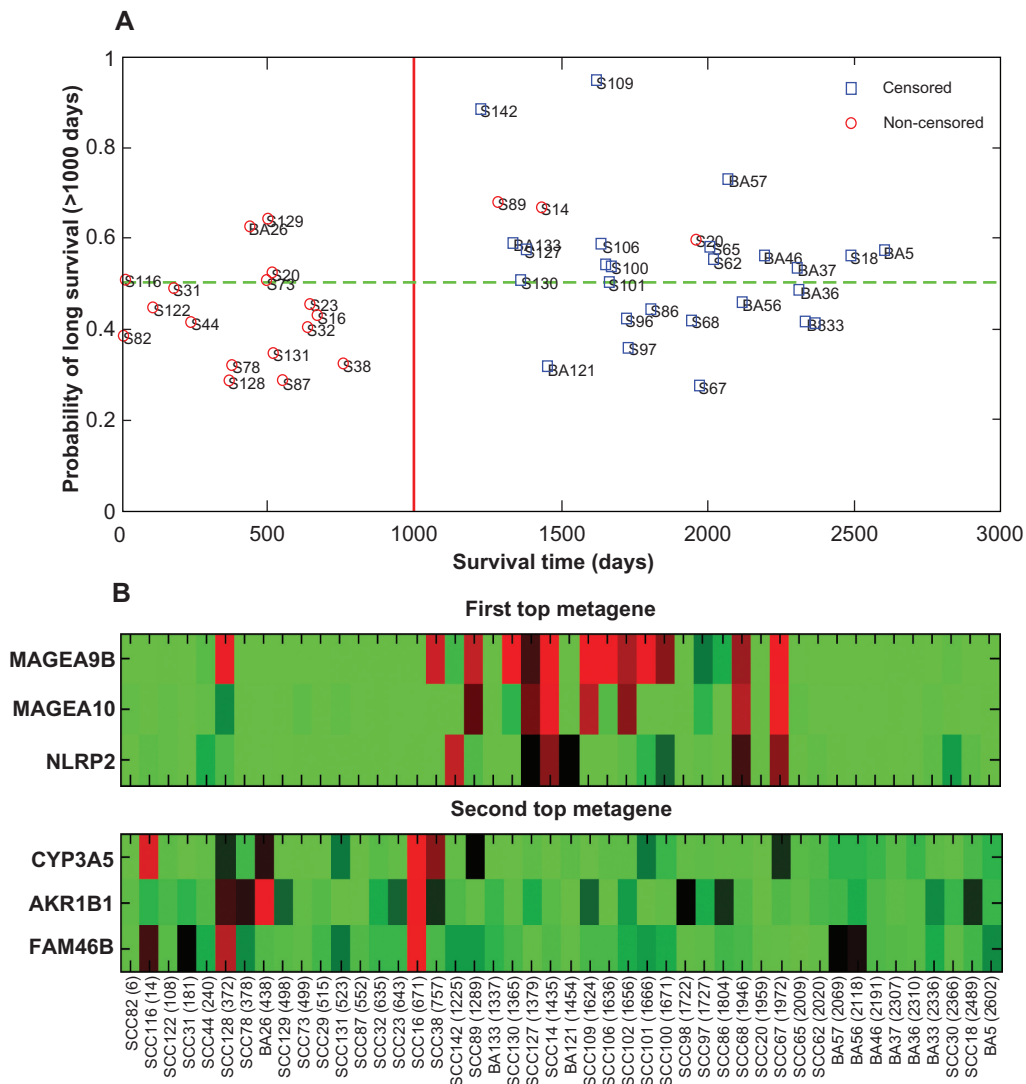


Figure 2. A) Leave-one-out cross-validation results demonstrating the ability of the model to classify new Ib patients into groups of low and high survival risk. The y-axis gives the posterior probabilities that a patient is classified into the long survival group (survival time > 1000 days) and the x-axis gives the observed survival time (including censored times). AC and BC patients are denoted by BA and SCC patients by S. Patients above the 0.5 posterior probability line (green dotted line) are classified into the long survival (low risk) group and patients below the line are classified into the short survival (high risk) group. **B)** A heatmap presentation of the two top metagenes most frequently found to be associated with the two risk groups in the Bayesian model. The genes included into the metagenes are listed on the y-axis and the patients sorted according to their survival times (days) are listed on the x-axis. High gene-expression values are shown with red color and low values with green ().

in the sense that their model considers both clinical data and gene data.

Predicting patient survival for stage Ib patients

The ability of the Bayesian model to correctly classify stage Ib patients into groups of short and long survival based on gene-expression profiles was evaluated using leave-one-out cross-validation and the ROC curve. The use of leave-one-out cross-validation provides an estimate of the ability of the model to predict survival of new patients (Fig. 2A). The cross-validation

classification error was 33% (15 out of 46 were misclassified) for all patients, 29% (5/17) for the short survival group (<1000 days), and 34% (10/29) for the long survival group (>1000 days).

Kaplan-Meier analysis of the model-predicted risk groups for Ib patients was performed. The difference between the survival rates in the high and low risk groups was significant ($P = 0.0007$, using the log-rank test) (Fig. 4). Further, according to the ROC analysis the overall performance of the classification method is positive AUC = 0.728 (P -value 0.001) (Fig. 3).

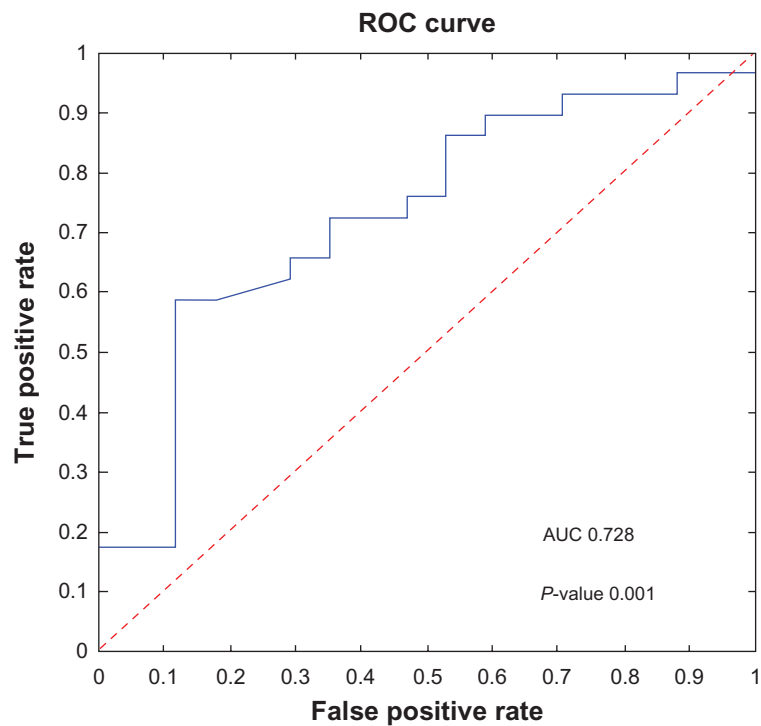


Figure 3. The empirical ROC curve (solid curve): The true positive rate plotted as a function of the false positive rate for different cut off values. Jumps in the curve correspond to changes in the classification outcome of the patients due to the use of different cut off values. For cut off value 0 all patients are classified into the long survival group, FTP = 0 and TPR = 0. In the other extreme, ie, cut off value 1 all patients are classified into the short survival group, FTP = 1 and FPR = 1. The area under the curve (AUC = 0.728, P -value 0.001) is an overall measure of the quality of the classification method. The dashed line (AUC = 0.5) is the expected ROC curve for a totally random classifier.

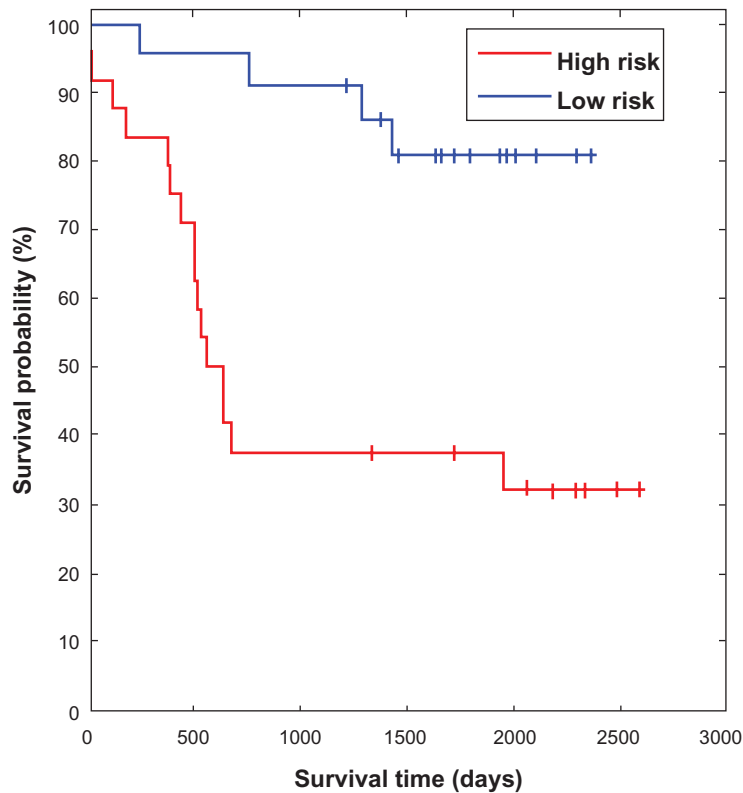


Figure 4. Kaplan-Meier plot of the survival probability in the high and low risk groups predicted by the Bayesian model. The high risk group (short survival) consisting of 24 patients and the low risk group (long survival) consisting of 22 patients. Vertical drops indicate deaths and ticks on the solid lines are censored survival times. The survival rates of the two groups are significantly different ($P = 0.0007$).



Conclusion

In this paper we applied two statistical approaches, one designed to identify alternating gene-expression patterns associated with tumor subtypes, and another designed to find genes associated with the risk for low survival in NSCLC patients. The approaches complement each other as the first approach provides information on gene-expression variation between different tumor subtypes, but does not address the important issue of different survival outcomes within patients classified into a single tumor stage. The second approach identifies genes associated with the risk for low survival in Ib stage patients which was the largest group of the patients. Our results definitely provide a possible strong reference for diagnosis/prognosis as the genes most up-regulated in pattern I and the genes most down regulated in pattern II may distinguish carcinogenesis progression for gene-expression pattern that corresponds to subtypes.

Acknowledgements

This research is supported by EU Lifespan—FP6 grant #36894, by Estonian Science Foundation grant ETF7859, by Targeted Financing from Estonian Government: SF0180142Cs08 and by OPENGENE grant FP7-REGPOT-2009-1. This research was also supported by the European Union through the European Regional Development Fund directed toward the Centre of Excellence in Genomics. This research work has been supported by the grant from The Special Government Funding allocated to Turku University Hospital.

Disclosure

This manuscript has been read and approved by all authors. This paper is unique and not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. Fry WA, Phillips JL, Menck HR. Ten-year survey of lung cancer treatment and survival in hospitals in the United States: a national cancer data base report. *Cancer*. Nov 1 1999;86(9):1867–76.
2. Hayat MJ, Howlader N, Reichman ME, Edwards BK. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *Oncologist*. Jan 2007; 12(1): 20–37.
3. Greene FL, Page DL, Fleming ID, et al. *AJCC cancer staging handbook*, 6th New York: Springer-Berlag; 2002:89–98.
4. Rami-Porta R, Chansky K, Goldstraw P. Updated lung cancer staging system. *Future Oncol*. Dec 2009;5(10):1545–53.
5. Goldstraw P, Crowley J, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol*. Aug 2007;2(8):706–14.
6. Boutros PC, Lau SK, Pintilie M, et al. Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci U S A*. Feb 24 2009;106(8):2824–8.
7. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med*. Aug 2008;14(8):822–7.
8. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*. Aug 2002;8(8): 816–24.
9. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res*. Aug 1 2006;66(15):7466–72.
10. Chen HY, Yu SL, Chen CH, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. Jan 4 2007;356(1):11–20.
11. Brambilla E, Travis WD, Colby TV, Corrin B, Shimamoto Y. The new World Health Organization classification of lung tumours. *Eur Respir J*. Dec 2001;18(6):1059–68.
12. Carr DT, Mountain CF. Staging lung cancer. In: Strauss M, ed. *Lung Cancer: Clinical Diagnosis and Treatment* New York: Grune and Stratton; 1977: 151–61.
13. Travis WD, Travis LB, Devesa SS. Lung cancer. *Cancer*. Jan 1 1995;75(1 Suppl): 191–202.
14. WHO. Histological typing of lung tumours. *Tumori*. Aug 1981;67(4):253–72.
15. Fleming ID, Cooper JS, Henson DE, et al. *AJCC Cancer staging manual 5th Philadelphia*: Lippincott-Raven; 1997.
16. Valk K, Voorder T, Kolde R, et al. Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers. *Oncology*.79(3–4): 283–92.
17. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit SR, Irizarry WH, editors. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: Springer; 2005:397–420.
18. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol*. 2007;8(1):R3.
19. Nogales-Cadenas R, Carmona-Saez P, Vazquez M, et al. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Res*. Jul 1 2009;37 (Web Server issue): W317–22.
20. Pittman J, Huang E, Dressman H, et al. Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. *Proc Natl Acad Sci U S A*. Jun 1 2004;101(22):8431–6.
21. Quackenbush J. Computational analysis of cDNA microarray data. *Nature Reviews*. 2001;6(2):418–28.
22. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993;88(422):669–79.
23. Hoti F, Sillanpaa MJ. Bayesian mapping of genotype x expression interactions in quantitative and qualitative traits. *Heredity*. Jul 2006;97(1):4–18.
24. Bae K, Mallick BK. Gene selection using a two-level hierarchical Bayesian model. *Bioinformatics*. Dec 12 2004;20(18):3423–30.
25. Kuner R, Muley T, Meister M, et al. Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer*. Jan 2009;63(1):32–8.
26. Saviozzi S, Cordero F, Lo Iacono M, Novello S, Scagliotti GV, Calogero RA. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC Cancer*. 2006; 6:200.
27. Wu M, Orta L, Gil J, Li G, Hu A, Burstein DE. Immunohistochemical detection of XIAP and p63 in adenomatous hyperplasia, atypical adenomatous hyperplasia, bronchioloalveolar carcinoma and well-differentiated adenocarcinoma. *Mod Pathol*. May 2008;21(5):553–8.



28. Kettunen E, Anttila S, Seppanen JK, et al. Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. *Cancer Genet Cytogenet.* Mar 2004;149(2): 98–106.
29. De Martino I, Visone R, Wierinckx A, et al. HMGA proteins up-regulate CCNB2 gene in mouse and human pituitary adenomas. *Cancer Res.* Mar 1 2009;69(5):1844–50.
30. Rohrbeck A, Neukirchen J, Roskopf M, et al. Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. *J Transl Med.* 2008;6:69.
31. Spira A, Beane JE, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med.* Mar 2007;13(3):361–6.
32. Franklin WA. RAGE in lung tumors. *Am J Respir Crit Care Med.* Jan 15 2007;175(2):106–7.
33. Lin J, Lin L, Thomas DG, et al. Melanoma-associated antigens in esophageal adenocarcinoma: identification of novel MAGE-A10 splice variants. *Clin Cancer Res.* Sep 1 2004;10(17):5708–16.
34. Pan JH, Han JX, Wu JM, Sheng LJ, Huang HN. CYP450 polymorphisms predict clinic outcomes to vinorelbine-based chemotherapy in patients with non-small-cell lung cancer. *Acta Oncol.* 2007;46(3):361–6.
35. Kolesar JM, Breunig A, Miller J, et al. CYP3A5*3 and CYP3A4*1B polymorphisms are associated and more frequent in NSCLC tumors than in normal volunteers. *Journal of Clinical Oncology, 2004 ASCO Annual Meeting Proceedings* July 15, 2004;22(14S):2016.
36. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med.* Aug 10 2006;355(6):570–80.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>