HLA2 and NEBL as novel prognostic biomarkers based on Rayleigh survival model in colorectal cancer

Altaf Khan (✉ ahkhan25@hotmail.com)

King Abdullah International Medical Research Center

Mohammed Azhar Aziz

King Abdullah International Medical Research Center

Ghada T Alatar

King Abdullah International Medical Research Center

Amal AlGhamdi

King Abdullah International Medical Research Center

Mohammed A Hussein

King Abdullah International Medical Research Center

Tanveer S Khatlani

King Abdullah International Medical Research Center

Research

Keywords: Indigenous Dataset, Exon Microarray Dataset, Maxstat Algorithm, Parametric and Semi-parametric Survival Models

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-301022/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background The primary focus for this study was investigation of prognostic biomarkers in colorectal cancer (CRC), since biomarkers are instrumental in clinical decision making and patient management as well as playing pivotal roles in precision medicine.

Methods We analyzed indigenous dataset from colorectal cancer patients. Exon microarray dataset was used and 135 genes were identified as novel candidate biomarkers. 135 genes were further split into two groups: low and high gene expression values via 'maxstat' algorithm, they were analyzed using Kaplan-Meier (K-M) analysis and univariate Cox model, and a set of 33 genes were identified as statistically significant ($p < 0.05$). Furthermore, using the 'VARCLUS' algorithms (a SAS software procedure) which is a useful tool for variable reduction, based on the divisive clustering techniques, a small subset of 5 genes were selected out of 33 significant genes as potential candidates to build survival models. Both parametric and semi-parametric survival models were utilized to assess whether these 5 genes could be used as prognostic biomarkers.

Results HHLA2 ($p < 0.01$) and NEBL ($p < 0.01$) genes emerged as potential biomarkers, based on the para- and semi-parametric models such as: Rayleigh, Exponential probability distributions, and Cox models and were also validated on independent datasets, apart from validation with qPCR test as well as with the cell lines patients data in the laboratory. A rigorous statistical evaluation for model's performance were done via Harrell's index, Brier score, Shoenfeld residual plots as well as with comparing several predictive model plots. We also made the comparison of Akaike and Bayesian Information (AIC and BIC) criteria as well as log-likelihood estimates. The Tukey-Anscombe plot and Quantile-Quantile plot as diagnostic tools were applied to validate the parametric survival models.

Conclusion Based on predictive models HHLA2 and NEBL novel biomarkers were found as statically significant.

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the latest manuscript can be downloaded and [accessed as a PDF](#).

Figures

Figure 1

An overall work-flow for building survival models is illustrated schematically.

biomarkers 33 genes

Figure 2

Based on univariate Cox-model and Kaplan-Meier survival plot, a subset of 33 genes were statistically significant and out of this subset of genes 16 genes are known to be associated with coloerctal cancer. IPA tool was utilized to make this graph.

Figure 3

In the left panel the violin plot for HHLA2 gene depicts the normal vs. tumor tissues gene expression values, based on the Wilcoxon test p-value shows high statistical significance ($p < 0.00001$), moreover shape (thicker or thinner) of the violin plots tell that the thicker part implies the values in that section of the violin has higher frequency, and the thinner part implies lower frequency. Whereas, on the right panel the violin plot for NEBL gene: normal vs. tumor tissues shows the expression values, applying the Wilcoxon p-value shows high statistical significance ($p < 0.00001$), moreover shape (thicker or thinner) of the violin plots tell that the thicker part implies the values in that section of the violin has higher frequency, and the thinner part implies lower frequency,

Figure 4

PCR validation: shows patients are down regulated with respect to fold change.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryCRC.pdf](#)