PanelomiX: a threshold-based algorithm to create panels of biomarkers

Xavier Robin[a], Natacha Turck[a], Alexandre Hainard[a], Natalia Tiberti[a], Frédérique Lisacek[b], Jean-Charles Sanchez[≈, a], Markus Müller[b]

[a] Translational Biomarker Group, Department of Human Protein Sciences, University of Geneva, Geneva, Switzerland.
[b] Proteome Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland.

April 2013

Abstract

In order to increase their predictive power, medical biomarkers can be combined into panels. However, the lack of ready-to-use tools generating interpretable results and implementing rigorous validation standards hampers the more widespread application of panels and their translation into clinical prac tice.

The computational toolbox we present here – PanelomiX – uses the iterative combination of biomark ers and thresholds (ICBT) method. This method combines biomarkers and clinical scores by selecting thresholds that provide optimal classification performance. To speed up the calculation for a large number of biomarkers, PanelomiX selects a subset of thresholds and parameters based on the random forest method. The panels' robustness and performance are analysed by cross-validation (CV) and re ceiver operating characteristic (ROC) analysis.

Using 8 biomarkers, we compared this method against classic combination procedures in the determi nation of outcome for 113 patients with an aneurysmal subarachnoid haemorrhage. The panel classified the patients better than the best single biomarker ($p < 0.005$) and compared favourably with other off-the-shelf classification methods.

In conclusion, the PanelomiX toolbox combines biomarkers and evaluates the performance of panels to classify patients better than single markers or other classifiers. The ICBT algorithm proved to be an efficient classifier, the results of which can easily be interpreted.

Citation: Robin X., Turck N., Hainard A., Tiberti T., Lisacek F., Sanchez J.-C. and Müller M. (2013). PanelomiX: a threshold-based algorithm to create panels of biomarkers. *Translational Proteomics* 1 (1) p. 57–64. DOI: 10.1016/j.trprot.2013.04.003.

Keywords: Biomarkers, Panel, Combination of biomarkers, Machine learning, Clinical study, Aneurysmal subarachnoid haemorrhage.

Abbreviations: ROC: receiver operating characteristic; AUC: area under the ROC curve; pAUC: partial AUC; CV: cross-validation; SE: sensitivity; SP: specificity; aSAH: aneurysmal subarachnoid haemorrhage; SVM: support vector machines.

Submitted: February 13th, 2013; Revised: April 24th, 2013; Accepted: April 26th, 2013.

Introduction

The translation of panels of biomarkers into clinical practice is principally obstructed by two critical factors (Robin et al. 2009). Firstly, methods and results can often be difficult to understand for non-experts; secondly, there is a general lack of robust validation steps, which are critical for the reproducibility of results given high biological variation.

To overcome the first issue, a combination method must produce clear and easily interpretable results, where patient classification can be understood in terms of the contribution of each individual biomarker. Medical practitioners have long been used to clinical scores, such as the Hoffer-Osmond test to diagnose schizophrenia (Hoffer & Osmond 1961; Kelm & Hoffer 1965), or the Ranson score (Ranson et al. 1974) for the prognosis and operative management of acute pancreatitis. These methods were recently applied to assess the probability of pulmonary embolism (Wicki et al. 2001) and acute pancreatitis (Imrie 2003). These types of scores have become popular because they are clear and easy to interpret, granting access

≈ Corresponding author. Centre Médical Universitaire, 1 rue Michel-Servet, CH-1211 Geneva 4, Switzerland. Tel.: +41 22 379 5486; e-mail address: Jean-Charles.Sanchez@unige.ch.

to the intermediate results of individual sub-tests. This is in contrast to black box classifiers, such as neural networks or support vector machines (SVM), which may display high accuracy, but which do not reveal the contribution of each individual marker directly. While black boxes are acceptable in specific applications, they may not always be suitable in expert systems for medical decision-making (Duch et al. 2004; Andrews et al. 1995; Baker 2005). In contrast, many methods present results in a user-friendly format referred to as "white boxes".

Combining biomarkers is an application of statistical learning. Over the years, this field has developed countless methods to tackle the task. Linear or logistic regression methods determine a factor, generally multiplicative, for each biomarker included in the panel. A straightforward interpretation of these factors is to see them as the "weights" of influence of the biomarkers. Methods based on decision trees also provide an easy interpretation, where one follows a sequence of binary splits. As long as a tree contains only a fairly limited number of such decisions (or branches), these are easy to track and to justify how a decision was reached. Decision trees are graphically expressive (see (Robin et al. 2009)) for easier understanding. Finally, in threshold-based methods, all biomarker tests are analysed at the same time (instead of sequentially), and the number of positive tests defines a score used for classification.

The second issue is the lack of a robust validation step. Panel validation requires an independent test set – preferably measured in a different laboratory – in order to compute the panel's true performance and avoid performance overestimation due to over-fitting the data during the learning process (Robin et al. 2009). If no independent set is available, computational methods such as cross-validation (CV) or bootstrapping allow the simulation of such sets (Hastie et al. 2003; Dziuda 2010).

Two useful and quite common performance measures are sensitivity (the proportion of positive patients correctly detected by the test) and specificity (the proportion of negative patients correctly rejected by the test), as they give clear estimates of how patients are classified (Robin et al. 2009). When no biomarker level cut-off is preferred or pre-defined, receiver operating characteristic (ROC) analysis can be performed to weight the trade-off between sensitivity and specificity (Hastie et al. 2003). The area under the ROC curve (AUC) is also a very common performance metric in medical decision-making (Pepe 2003), bioinformatics (Sonego et al. 2008) and statistical learning (Fawcett 2006). An important and often neglected step is the panel's performance comparison against that of single biomarkers. A fair evaluation would process the panel and single biomarkers with the same tools (sensitivity and specificity or AUC) on the same independent test set or with the same CV procedure (Robin et al. 2009). Then performance could be compared either with McNemar's test (for sensitivity or specificity) or using ROC curves.

The method we propose here is called PanelomiX. In this paper, we use threshold as the base of decisions. In threshold-based combinations, thresholds are often chosen in a univariate manner. For example, Ranson *et al.* (Ranson et al. 1974) selected convenient prognostic sign cut-off values outside the range of the mean plus or minus one standard deviation; Morrow and Braunwald (Morrow & Braunwald 2003) chose the 99th percentile of the control distribution; Sabatine et al. (Sabatine et al. 2002) used the cut-offs described in the literature. In contrast, Reynolds *et al.* (Reynolds et al. 2003) adopted a multivariate approach and tested many thresholds by 10% increments. This approach takes into account the interaction that may arise when biomarkers are combined.

PanelomiX can combine biomarkers (molecule levels, clinical scores, etc.) in a multivariate manner. Therefore we developed an exhaustive search algorithm to select the threshold, and called it iterative combination of biomarkers and thresholds (ICBT). To minimize execution times, we developed several approaches to reduce complexity and hence increase search speed. As it has been shown to be an efficient feature selection method (Dziuda 2010), we used random forest (Breiman 2001; Liaw & Wiener 2002) as a filtering method to reduce both the number of biomarkers and thresholds that account for the search space size. Random forest builds a large number of decision trees that are made slightly different by bootstrapping. In the end, the classification is the average prediction of all trees.

PanelomiX has already been applied to predict the outcome of an aneurysmal subarachnoid haemorrhage (aSAH) (Turck et al. 2010) and to assess the progression of human African trypanosomiasis (Hainard et al. 2009). Below, we demonstrate the PanelomiX methodology and performance, using 8 parameters for the determination of outcome for patients with an aSAH.

Methods

Iterative combination of biomarkers and thresholds (ICBT)

Combining biomarkers

The approach adopted here is based on the ICBT method. A threshold is defined for each biomarker by an optimization procedure defined in the following sections. A patient's score is the number of biomarkers exceeding their threshold values.

We can write this as:

$$S_p = \sum_{i=1}^{n} I\left(X_{ip} > T_i\right), \quad \text{Equation 1}$$

where S_p is the score for patient p, n is the number of biomarkers, X_{ip} is the concentration of the ith biomarker in patient p, T_i is the threshold for the ith biomarker, and I(x) is an indicator function which takes the value of 1 for x = true and 0 otherwise.

If biomarker concentrations are higher in the control than in the disease group, then they are multiplied by -1 before applying the previous formula.

To classify a patient, a threshold on the S_p score is required and defined as T_s. Patients with a score $S_p \geq T_s$ are positive; negative otherwise.

Selecting the thresholds

The list of thresholds tested in the ICBT search must be kept short to limit computation time. Candidate thresholds are selected as local extremums of the ROC curve, computed with pROC (Robin et al. 2011). A local extremum is defined as a point of local maximal distance to the diagonal line. To construct the ROC curve we sort the list of biomarker values, resulting in a list of increasing specificity (SP) and decreasing sensitivity (SE). The threshold value T_i is a local extremum if SP[i] ≥ SP[i-1] and SE[i] ≥ SE[i+1]. Thresholds that are not local extremums will not lead to better classification. Several thresholds are usually extremums on a ROC curve.

Optimizing the panel

The combinatorial complexity of testing all combinations of biomarkers and threshold values with ICBT can be calculated. Given n biomarkers, and panels with up to m biomarkers, the number C of biomarker combinations to test, is given by:

$$C = \sum_{i=1}^{m} \binom{n}{i} = \sum_{i=1}^{m} \frac{n!}{i!(n-i)!} \quad \text{Equation 2}$$

If there are t thresholds per biomarker, formula 3 gives the total number I of threshold combinations to test:

$$I = \sum_{i=1}^{m} \left(\frac{n!}{i!(n-i)!} t^i \right) \quad \text{Equation 3}$$

In addition, all possible T_s from 1 to n-1 are considered.

In a typical setup, one would test combinations of 5 or less out of 10 biomarkers, with 15 thresholds per biomarker. This corresponds to 637 possible biomarker combinations to test. The total number of possible combinations of thresholds and biomarkers comes to 202 409 025, which is still manageable using current desktop computers.

In most real world applications, however, each biomarker will have a different number of thresholds. If T is a vector containing the number of thresholds of all biomarkers in combination j, a more precise estimate is given by:

$$I = \sum_{j=1}^{C} \left(\prod T_j \right) \quad \text{Equation 4}$$

Pre-filtering

When computational time becomes too long, an additional step is necessary to reduce the number of biomarkers and thresholds. From the N initial biomarkers, P biomarkers are selected (with $P < N$), each associated with Q cut-offs. . In PanelomiX, random forest (Breiman 2001; Liaw & Wiener 2002) is

employed as a multivariate filter (Dziuda 2010). The trees created during the process are analysed to deduce the most frequent biomarkers and thresholds that give the most interesting combinations.

We proceed by stepwise elimination. First, a random forest with all the N biomarkers is created. The frequency each biomarker appears in tree branches is extracted and the N-1 biomarkers occurring most often are kept to build the next random forest. These two steps are repeated until the target number of P biomarkers is reached. Finally, a last random forest is computed with P remaining biomarkers to determine the Q thresholds occurring most frequently for each marker. As each tree of the random forest is computed from a different set of patients, the cut-offs will differ slightly between the decision trees of the forest. To be more informative, the thresholds are therefore mapped to the original ones using Euclidean distance. Thresholds are then sorted by frequency and the Q first thresholds of each biomarker are selected for an exhaustive search.

Code optimization

At the programming level, the ICBT search was optimised to run faster. First, it was implemented in the compiled programming language Java, which typically runs much faster than interpreted languages such as R, Perl or Python. Efficient implementation was achieved by minimizing the creation of objects, using explicit programmatic loops instead of recursion and multithreading.

Biomarkers with missing values are ignored. Missing value imputations must be performed before submitting the data to PanelomiX (see (Aittokallio 2010) for an in-depth review of this topic).

Cross-validation

Cross-validation (CV) is a simple and widely used computational method to assess a classification model's performance and robustness (Hastie et al. 2003; Robin et al. 2009). PanelomiX features a CV procedure for panel verification (Hastie et al. 2003). Its primary goal is to test panel performance in an unbiased manner and to produce graphical diagnostic plots for evaluating consistency and robustness. After CV, ROC analyses were performed on the individual biomarkers and the panel, and several plots were generated to assess the quality of the data.

A standard, k-fold cross-validation (CV) scheme was used to compare the different models generated. To avoid model-to-model scoring differences and make predictions comparable between the CV steps, which may produce panels of different lengths with different Ts, the prediction is centred as follows:

$$Y_p = S_p - T_s \qquad \text{Equation 5}$$

$$Z_p(Y_p) = \begin{cases} Y_p / T_s, Y_p < 0 \\ Y_p / (n - T_s), Y_p > 0 \end{cases} \qquad \text{Equation 6}$$

As a result, the centred vector Z of patient scores is in the [-1;+1] interval and T_s =0.

ROC curves

We perform ROC analysis of the curves of both the individual biomarkers and the panels using the pROC tool (Robin et al. 2011) in R (R Development Core Team 2008). Three tables are generated showing AUC, sensitivity, and specificity, all with confidence intervals. The first table reports the ROC performance of single biomarkers and their best univariate thresholds; the second table shows the comparison of the panel with the best individual biomarker (analyzed as a panel composed of 1 biomarker, to be comparable with the panels); and the third table compares the ICBT panel with other classic combination methods. Comparisons between two AUCs are performed using DeLong's test (DeLong et al. 1988) and between two pAUCs using the bootstrap test (Robin et al. 2011) with 10 000 stratified replicates. The ROC curves of the CV are built as the mean of centred predictions over the k CV folds. For the CV of the individual biomarkers, the ICBT algorithm is applied with n = 1 and no other modification.

Availability

Users can access a password-protected server implementing the algorithms described in this article from the following website: http://www.panelomix.net.

Case study

Patients

The PanelomiX methodology was applied to a previously published data set of 113 patients with an aSAH. The goal was to identify patients at risk of a poor outcome six months after an aSAH – those who would require specific healthcare management. Detailed results of the study are reported in (Turck et al. 2010). We will only outline the features relevant to panel analysis here.

Panel analysis

As described above, panels were generated with five proteins (H-FABP, S100β, Troponin I, NKDA and UFD-1) and three clinical factors (WFNS, Modified Fisher score and age). A ten-fold CV was carried out to assess the performance of the biomarkers, the panels and their stability.

Comparison with standard methods

The results obtained with PanelomiX were compared with other methods: logistic regression with the glm package and step-wise elimination functions; support vector machines (SVM) using the kernlab package (Karatzoglou et al. 2004) (nu-regression with linear kernel); and recursive partitioning decision trees using the rpart package (Therneau & Atkinson 1997; Therneau et al. 2012). To be consistent with the PanelomiX method, both SVM and decision tree feature sets were determined using an exhaustive search of all possible combinations. Additionally, the predictions were centred as described above.

ROC sample size computation

The sample size required for a statistically significant comparison of two ROC curves was calculated according to Obuchowski and McClish (Obuchowski & McClish 1997), where variances and covariances of the ROC curves were computed using bootstrapping (Efron & Tibshirani 1993).

Results and discussion

Training the panels

The PanelomiX methodology was applied to the 113-patient cohort of the aneurysmal subarachnoid haemorrhage study (Turck et al. 2010) in order to define the combination of 8 biomarkers with the best classification accuracy. Using the whole cohort as a training set, but without CV, a panel containing 8 biomarkers (i.e. the 5 proteins and the 3 clinical parameters) was found using the thresholds given in 1. The panel's performance was evaluated using two methods: threshold sensitivity and specificity, and area under the ROC curve (AUC). On the training set this panel showed 95% sensitivity and 90% specificity, corresponding to an AUC of 95%.

Biomarker	H-FABP	S100b	Troponin I	NDKA	UFD-1	WFNS	Age	Fisher Score
Threshold	1.11	0.51	2.33	11.08	271.48	1.5	72.5	2.5
Unit	µg/l	µg/l	µg/l	µg/l	µg/l	N/A	Years	N/A

Table 1: Biomarkers and thresholds in the panel

Cross-validation

Tenfold CV was repeated 10 times with 10 random selections of the folds. The four plots that allowed us to evaluate the stability of the panel with CV are shown in Figure 1.

- The marker selection frequency plot shows the frequency of selection of each biomarker variable in the panels trained in k CV folds. A biomarker with a 100% frequency is selected in all panels; the frequency is weighted. If one step of the CV yields several panels, then each of them contributes less to the final frequency compared to panels which were unique in a CV fold. Figure 1A shows that all eight biomarkers selected in the training panel are selected between 88% (Fisher score) and 100% (NDKA, H-FABP, S100b, WFNS) of the CV panels.
- The panel size frequency plot displays the number of biomarkers in the panels, weighted as described above. Figure 1B shows that 69% of the CV panels contained 8 biomarkers. In 27% of

the CV panels 7 biomarkers were selected, and in the 4% remaining only 6 biomarkers were selected. No panels containing 5 or fewer biomarkers were generated during the CV procedure.

- The panel Ts frequency plot shows the score Ts, determining how many biomarkers must be positive in a patient for the panel to be positive, weighted as described above. In Figure 1C, Ts = 3 in 25% of the panels, Ts = 5 in 4% and Ts = 4 in the rest of the cases.
- The threshold stability plot represents biomarkers on the x-axis and thresholds (as a rank, not an absolute value) of all panels found in the CV on the y-axis. Each panel corresponds to a line joining its constituting set of biomarkers and thresholds. Figure 1D shows that S100b had a very stable threshold, unlike NDKA or UFD-1 that showed a larger variation. For H-FABP, 3 clusters appeared, corresponding to thresholds of 0.61µg/l (rank 22), 1.11µg/l (rank 33) and 4.51µg/l (rank 84). This indicates that the H-FABP cut-off at 5.9µg/l, found in the training panel, is not as robust as the cut-off at 0.51µg/l found for S100b.

Performance evaluation

A ROC analysis was performed as described in the previous section (Figure 1). The panel found using the training set was plotted together with that found using CV and the separate biomarkers (see next section). Using CV, panels displayed 65.9% sensitivity and 88.9% specificity, corresponding to an AUC of 88.6%.

Figure 3 shows the performance of PanelomiX on the training set and using CV for panels of different sizes. Using CV, panels with 7 biomarkers are optimal, with an AUC (88.8%) slightly higher than panels of 8 (88.6%). However, the difference is minimal and it is difficult to determine the significance of this change. This indicates that the level of over-fitting induced by ICBT is low and that classification with panels is an improvement on single biomarkers.

Comparison with single biomarkers

Figure 3 shows that individual biomarkers are slightly over-fitted and display a lower AUC using CV (71%) than on the training sample (73%). To perform a fair comparison, PanelomiX compared both panel and single biomarkers under CV. To that end, we used the ICBT algorithm where the threshold is chosen on the training set, and applied to the test set.

The two best biomarkers, H-FABP and WFNS, are plotted with ICBT in Figure 1. The CV results (dotted lines) show that panels of 8 biomarkers, with an AUC of 89%, are superior to the individual biomarkers with AUCs of 76% (p = 0.003) for WFNS and 68% (p = 1.5×10-6) for H-FABP.

Comparison with established methods

PanelomiX was compared with three established methods of biomarker analysis: logistic regression, SVM and decision trees (recursive partitioning). The results are shown in Figure 4. PanelomiX displayed the best AUC (89%), slightly but not significantly higher than SVM (82%, p=0.20) and logistic regression (81%, p=0.13). Only recursive partitioning decision trees had a significantly lower AUC of 77% (p=0.03). Compared with SVM, PanelomiX gives results with a very similar classification performance, but in a way that is easier to interpret.

Evaluation of random forest pre-processing

Classification performance was assessed both with and without the initial pre-processing step using random forest. The results are shown in Figure 5. Pre-filtering made no difference in classification efficiency using one biomarker. However, as we tested panels of 2 to 6 biomarkers, it consistently led to decreased AUC. Among the possible explanations for this difference, is that the diagnostic plots (data not shown) indicated a selection of panels with fewer biomarkers when features were selected with random forest; this suggests that the tree-based feature selection is not optimal when combined with a threshold-based classification. With 7 and 8 biomarkers, the effect was reversed and the classification was even slightly improved when all biomarkers were selected. These results suggest that the pre-processing with random forest should be applied with care, and that a few more features than simply the target number should be kept in mind.

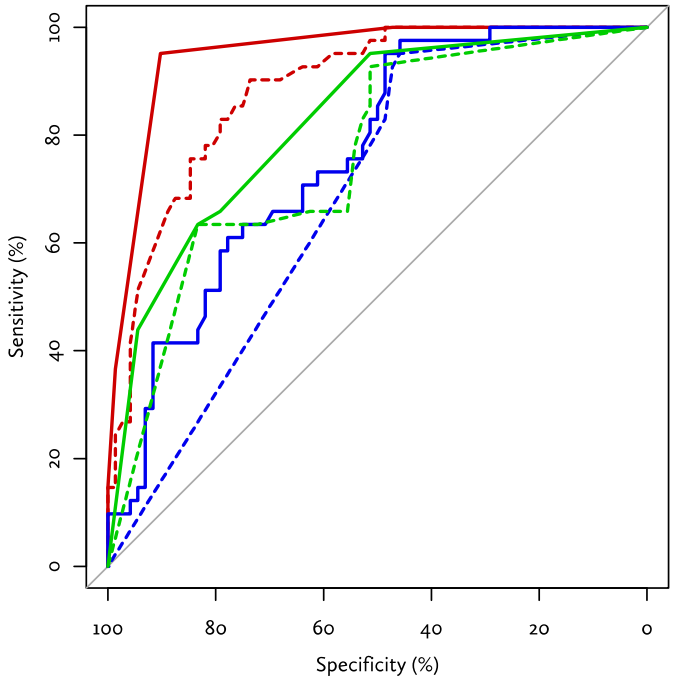

Figure 1: ROC curves. Solid lines represent the performance on the train ing set, dotted lines the cross-validation. Red: panel of 8 markers; green: WFNS; blue: H-FABP.

Computation time

As stated earlier, all the combinations of all 8 biomarkers and thresholds can be tested. 2 shows the processing time to train a single panel and to perform 10 ten-fold CVs. The CV of panels of up to 8 biomarkers took slightly less than 6 days to complete on a 4-core machine. Feature selection with random forest made computation 70 times faster with 8 markers, making it possible to test panels much larger than 8 biomarkers when necessary.

Panel size, *n*	1	2	3	4	5	6	7	8
Training only	0.25 s	0.34 s	1.2 s	8.2 s	1.6 min	11 min	49 min	2.0 h
Cross-validation	25 s	32 s	2.0 min	9.6 min	1.5 h	15 h	2.0 d	4.4 d
Cross-validation with random forest	1.1 min	1.4 min	1.6 min	1.7 min	3.3 min	9.9 min	25 min	1.6 h

Table 2: Execution time for increasingly large panels on an Intel Core 2 Quad CPU Q9550 at 2.83GHz processor. Table shows a simple training, and CV (N=10, K=10).

Conclusions

In this paper, we proposed an algorithmic solution for combining several biomarkers into a panel using the ICBT method based on an iterative combination of biomarkers and thresholds. We demonstrated that the definition of an optimal panel through exhaustive search is feasible with current computers. Unlike the 10% increments adopted by Reynolds et al. (Reynolds et al. 2003), the set of cut-offs to be tested is selected from the local extremum points on the ROC curve. This guarantees an optimal classification, and is better suited to the non-normally distributed data commonly found in clinical studies, where the last increments may not be as significant as the first ones. Panels created with this methodology are robust and easy to understand, even to non-mathematicians. They provide efficient classification when compared with classic methods. We also proposed an approach to reduce the complexity and increase the speed of the search for larger data sets with random forest, efficiently limiting information loss.

Finally, we showed how to apply the method to answer a real clinical question that was the outcome prediction for 113 patients following an aneurysmal subarachnoid haemorrhage. Further validation studies will be necessary to show whether the ICBT algorithm performs better than classic methods. We

could nonetheless show that the classification power of the resulting panel is superior to that of single biomarkers. However, to be strictly validated these findings need to be replicated in larger, independent cohorts of patients. This step is often omitted in biomarker research. This omission turns out to be even more critical with panels of biomarkers which are more prone to over-fitting the data. Despite the application of cross-validation, proper validation studies with external cohorts of patients will be required to strengthen the conclusions reached through tools such as PanelomiX before the validity of these results will be trusted by researchers.

The study analyzes 8 biomarkers, however they were all discovered using univariate approaches and some of them were relatively highly correlated (Turck et al. 2010). Multivariate discovery approaches (Erler & Linding 2010) are beyond the scope of this paper, but they could potentially highlight more interesting combinations of biomarkers.

In the clinics, a panel of biomarkers would be employed in a very similar way than a single biomarker currently is. The only difference is that several measurements must be performed to reach a result. This has been demonstrated as feasible using point-of-care test (POCT) units (Macdonald & Nagree 2008; Saenger & Christenson 2010). However, POCT often lack good biomarker targets, a fact PanelomiX could hopefully help solving.

Future prospects include the application of this workflow to data sets with more biomarkers, for instance coming from gene or protein microarrays or Single Reaction Monitoring experiments. It could also potentially be applied to the discovery of new biomarkers displaying higher classification performance when combination with other biomarkers.

Acknowledgements

This work was partially funded by Proteome Science PLC.

Authors' contributions

XR carried out the programming and software design, and drafted the manuscript. NTu, AH, NTi provided data and biological knowledge, and tested and critically reviewed the software and the manuscript. FL helped to draft and critically improve the manuscript. JCS conceived the biomarker study, participated in its design and coordination, and helped to draft the manuscript. MM participated in the design and coordination of the bioinformatics part of the study, participated in the programming and software design, and helped to draft the manuscript. All authors read and approved the final manuscript.

References

Aittokallio T., (2010). Dealing with missing values in large-scale studies: microarray data imputation and beyond. *Briefings in Bioinformatics*, 11 (2), p. 253 -264. DOI: 10.1093/bib/bbp059.

Andrews R., Diederich J. & Tickle A. B., (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems*, 8 (6), p. 373-389. DOI: 10.1016/0950-7051(96)81920-4.

Baker M., (2005). In biomarkers we trust? *Nat Biotech*, 23 (3), p. 297-304. DOI: 10.1038/nbt0305-297.

Breiman L., (2001). Random Forests. *Machine Learning*, 45 (1), p. 5-32. DOI: 10.1023/A:1010933404324.

DeLong E. R., DeLong D. M. & Clarke-Pearson D. L., (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44 (3), p. 837-845.

Duch W., Setiono R. & Zurada J. M., (2004). Computational intelligence methods for rule-based data understanding. *Proceedings of the IEEE*, 92 (5), p. 771- 805. DOI: 10.1109/JPROC.2004.826605.

Dziuda D. M., (2010). *Data Mining for Genomics and Proteomics: Analysis of Gene and Protein Expression Data*, John Wiley & Sons.

Efron B. & Tibshirani R. J., (1993). *An Introduction to the Bootstrap* Chapman & Hall., New York, London.

Erler J. T. & Linding R., (2010). Network-based drugs and biomarkers. *The Journal of pathology*, 220 (2), p. 290–296.

Fawcett T., (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27 (8), p. 861-874. DOI: 10.1016/j.patrec.2005.10.010.

Hainard A., Tiberti N., Robin X., et al., (2009). A Combined CXCL10, CXCL8 and H-FABP Panel for the Staging of Human African Trypanosomiasis Patients. *PLoS Neglected Tropical Diseases*, 3 (6), p. e459. DOI: 10.1371/journal.pntd.0000459.

Hastie T., Tibshirani R. & Friedman J., (2003). *Elements of Statistical Learning: data mining, inference, and prediction* Springer-Verlag., New York.

Hoffer A. & Osmond H., (1961). A card sorting test helpful in making psychiatric diagnosis. *Journal of Neuropsychiatry*, 2, p. 306-330.

Imrie C. W., (2003). Prognostic indicators in acute pancreatitis. *Canadian Journal of Gastroenterol*, 17 (5), p. 325–328.

Karatzoglou A., Smola A., Hornik K., et al., (2004). kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11 (9), p. 1–20.

Kelm H. & Hoffer A., (1965). A revised score for the Hoffer-Osmond diagnostic test. *Diseases of the Nervous System*, 26 (12), p. 790-1.

Liaw A. & Wiener M., (2002). Classification and Regression by randomForest. *R News*, 2 (3), p. 18-22.

Macdonald S. P. J. & Nagree Y., (2008). Rapid risk stratification in suspected acute coronary syndrome using serial multiple cardiac biomarkers: a pilot study. *Emergency medicine Australasia: EMA*, 20 (5), p. 403-409. DOI: 10.1111/j.1742-6723.2008.01116.x.

Morrow D. A. & Braunwald E., (2003). Future of Biomarkers in Acute Coronary Syndromes Moving Toward a Multimarker Strategy. *Circulation*, 108 (3), p. 250-252. DOI: 10.1161/01.CIR.0000078080.37974.D2.

Morrow D. A., Rifai N., Antman E. M., et al., (1998). C-Reactive Protein Is a Potent Predictor of Mortality Independently of and in Combination With Troponin T in Acute Coronary Syndromes: A TIMI 11A Substudy. *Journal of the American College of Cardiology*, 31 (7), p. 1460-1465. DOI: 10.1016/S0735-1097(98)00136-3.

Obuchowski N. A. & McClish D. K., (1997). Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Statistics in Medicine*, 16 (13), p. 1529-1542. DOI: 10.1002/(SICI)1097-0258(19970715)16:13<1529::AID-SIM565>3.0.CO;2-H.

Pepe M. S., (2003). *The statistical evaluation of medical tests for classification and prediction*, Oxford: Oxford University Press.

Ranson J. H., Rifkind K. M., Roses D. F., et al., (1974). Prognostic signs and the role of operative management in acute pancreatitis. *Surgery, Gynecology & Obstetrics*, 139 (1), p. 69-81.

R Development Core Team, (2008). *R: A Language and Environment for Statistical Computing*, Vienna, Austria.

Reynolds M. A., Kirchick H. J., Dahlen J. R., et al., (2003). Early Biomarkers of Stroke. *Clinical Chemistry*, 49 (10), p. 1733-1739. DOI: 10.1373/49.10.1733.

Robin X., Turck N., Hainard A., et al., (2009). Bioinformatics for protein biomarker panel classification: What is needed to bring biomarker panels into in vitro diagnostics? *Expert Review of Proteomics*, 6 (6), p. 675-689. DOI: 10.1586/EPR.09.83.

Robin X., Turck N., Hainard A., et al., (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, p. 77. DOI: 10.1186/1471-2105-12-77.

Sabatine M. S., Morrow D. A., De Lemos J. A., et al., (2002). Multimarker Approach to Risk Stratification in Non-ST Elevation Acute Coronary Syndromes Simultaneous Assessment of Troponin I, C-Reactive Protein, and B-Type Natriuretic Peptide. *Circulation*, 105 (15), p. 1760-1763. DOI: 10.1161/01.CIR.0000015464.18023.0A.

Saenger A. K. & Christenson R. H., (2010). Stroke Biomarkers: Progress and Challenges for Diagnosis, Prognosis, Differentiation, and Treatment. *Clinical Chemistry*, 56 (1), p. 21-33. DOI: 10.1373/clinchem.2009.133801.

Sonego P., Kocsor A. & Pongor S., (2008). ROC analysis: applications to the classification of biological sequences and 3D structures. *Briefings in Bioinformatics*, 9 (3), p. 198-209. DOI: 10.1093/bib/bbm064.

Therneau T. M., Atkinson B. & Ripley B., (2012). *rpart: Recursive Partitioning*, 3.1.52.

Therneau T. M. & Atkinson E. J., (1997). *An Introduction to Recursive Partitioning Using the RPART Routines*, Rochester, MN: Mayo Clinic.

Turck N., Vutskits L., Sanchez-Pena P., et al., (2010). A multiparameter panel method for outcome prediction following aneurysmal subarachnoid hemorrhage. *Intensive Care Medicine*, 36 (1), p. 107–115. DOI: 10.1007/s00134-009-1641-y.

Wicki J., Perneger T. V., Junod A. F., et al., (2001). Assessing Clinical Probability of Pulmonary Embolism in the Emergency Ward: A Simple Score. *Archives of Internal Medicine*, 161 (1), p. 92-97. DOI: 10.1001/archinte.161.1.92.