Detection of Prostate Cancer using Voltammetric Electronic Tongue.

Lluís Pascual [a], *Inmaculada Campos* [a], *José-Luis Vivancos* [a,b,c,d], *Guillermo Quintás* [e,f], *Alba Loras* [g], *M.Carmen Martínez-Bisbal* [a,c], *Ramón Martínez-Máñez* [a,c,d,h], *Francisco Boronat* [g], *José Luis Ruiz-Cerdà* [d,g]

Supporting Information

Data Analysis

In Prediction Results, Sensitivity (Sn) was defined as the fraction of the predicted samples that are correct amongst those predicted: Sn = Tp/(Tp +Fn), where Tp are True Positives and Fn are False Negatives. Specificity (Sp) was defined as the fraction of the predicted that are correct amongst those predicted: Sp = Tp/(Tp +Fp).

Pulse waveform

The pulse sequence for noble electrodes (Ir, Rh, Pt, Au) was: 0, 50, 0, 200, 0, 400, 0, 600, 0, 800, 0, 1000, 0, 800, 0, 600, 0, 400, 0, 200, 0, -200, 0, -400, 0, -600,0, -800, 0, -1000, 0, -800, 0, -600, 0, -400, 0, -200, 0, 0, 0, 0 (mV). In the case of non-noble electrodes (Ag, Co and Cu) the pulse sequence was: 0, 50, 0, 100, 0, 200, 0, 300, 0, 400, 0, 500, 0, 400,

[a] Instituto Interuniversitario de Reconocimiento Molecular y Desarrollo Tecnológico (IDM), Unidad Mixta Universitat Politècnica de València - Universitat de València.
[b] Departamento de Proyectos de Ingeniería, Universitat Politècnica de València, Camino de Vera s/n., 46022 Valencia, Spain.
[c] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BNN), Spain.
[d] Unidad Mixta de Investigación en Nanomedicina y Sensores, Universitat Politécnica de València – IIS La Fe de Valencia
[e] Leitat Technological Center, Safety and sutainability Division, Valencia, Spain
[f] Unidad Analítica, Instituto de Investigación Sanitaria Hospital La Fe, Valencia, Spain
[g] Servicio de Urología, Hospital Universitario y Politécnico La Fe, Valencia, Spain
[h] Departamento de Química, Universitat Politècnica de València, Camino de Vera s/n., 46022 Valencia, Spain

0, 300, 0, 200, 0, 100, 0, -100, 0, -200, 0, -300,0, -400, 0, -500, 0, -400, 0, -300, 0, -200, 0, -100, 0, 0, 0, 0 (mV).

Models Analysed

After evaluation of the predictive performance of different electrode combinations, three of them were selected using the cross-validated classification errors as selection criterion, namely C) Ir, Au and Ag Electrodes, D) Ir, Pt, Au and Ag Electrodes, E) Ir, Rh, Pt, Au and Ag Electrodes.

PCA

Initially, a principal component analysis (PCA) model was built using the whole data set and autoscaling as data pretreatment for data collected using the Ir, Au and Ag electrodes. From the scores plot of the first versus the second principal components and from the Q-residual vs Hotelling T^2, three samples were classified as outliers (samples 50, 72 and 98).

Same procedure was done for Model D). From the scores plot of the first versus the second principal components and from the Q-residual vs Hotelling T^2 three samples included on the whole data set were classified as outliers (samples 50, 72 and 98).

Same procedure was done for Model E). From the scores plot of the first versus the second principal components and from the Q-residual vs Hotelling T^2 three samples included on the whole data set were classified as outliers (samples 50, 72 and 98).

PLSDA Models

Supervised discriminant analysis was performed using partial least squares (PLSDA) and a maximum number of 7 latent variables (LVs). Selection of the number of LVs

was carried out using as figure of merit the mean number of misclassified samples after 12 iterations of a 8-fold random cross-validation (LV=min{NMC=FP+FN}). The X-block was autoscaled and the y vector containing class labels (i.e. -1 and +1 for control and PCa samples, respectively) was mean centered. The residual Q and the Hotelling's T^2 statistics were also used for outlier detection.

Split into Calibration and Validation using Kennard-Stone method at 66%. The method begins by finding the two samples that are farthest apart using geometric distance. To add another sample to the selection set the algorithm selects from the remaining samples that one which has the greatest separation distance from the selected samples. The separation distance of a candidate sample from the selected set is the distance from the candidate to its closest selected sample. This most separated sample is then added to the selection set and the process is repeated until the required number of samples, k, have been added to the selection set. In practice this produces a very uniformly distributed network of selected points over the data set and includes samples along the boundary of the dataset. The method performs efficiently because it calculates the inter-sample distances matrix only once [S1]. This resulted in a sensitivity of 91% and a specificity of 83%.

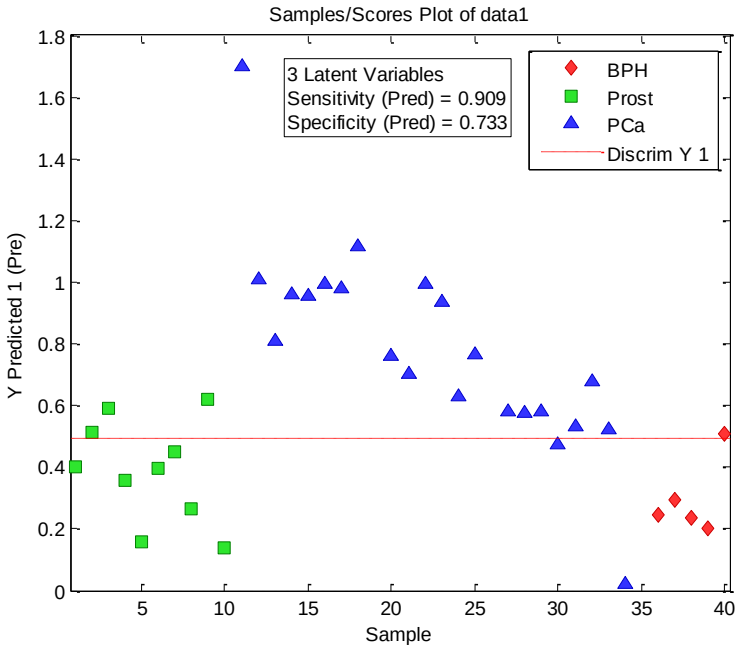

Figure S1. Prediction of validation dataset using 3 LVs using Ir, Au and Ag Electrodes. Model has been created using 2 classes (PCa and Control) only for visualization purpose samples from patients after radical prostatectomy (Prost) and patients diagnosed benign prostatic hyperplasia (BPH) are shown.

Table S1. Confusion table build using results obtained by PLSDA using Ir, Au and Ag Electrodes and 3 LVs. 74 samples were used for calibration and 37 samples for model validation. Wilcoxon =0.001 and 0.000 for Cross-Validated.

	validation set	
	PCa	**Control**
Predicted as PCa	**20**	**4**
Predicted as control	**2**	**11**
	22	**15**

Model D) Ir, Pt, Au and Ag Electrodes,

Supervised discriminant analysis was performed using PLSDA and a maximum number of 7 LVs. Selection of the number of LVs was carried out using the selection criterion described for Model C. Besides, the same procedure (i.e. Kennard-Stone method at 66%) was employed for the selection of the calibration and validation subsets. This resulted in a sensitivity of 100% and a specificity of 83%.

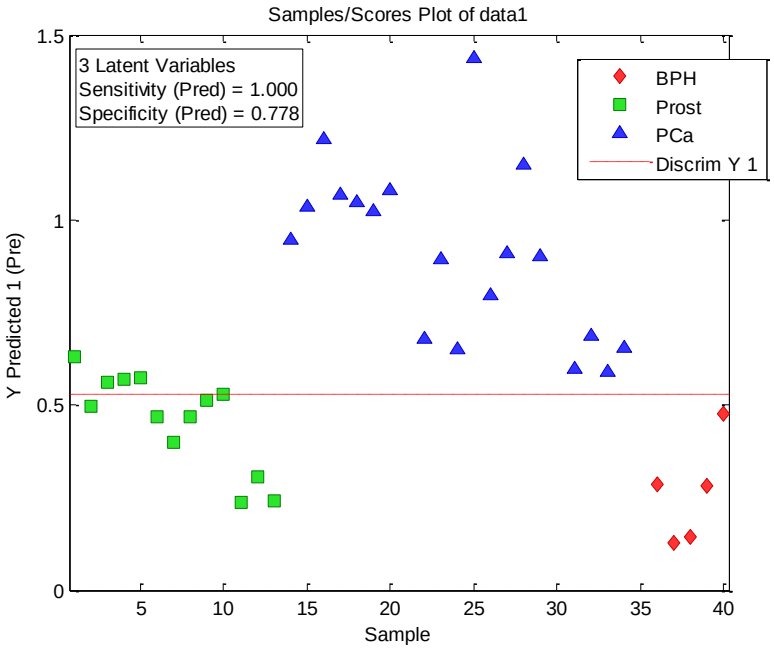

Figure S2. Prediction of validation dataset using 3 LVs using Ir, Au and Ag Electrodes. Model has been created using 2 classes (PCa and Control) only for visualization purpose samples from patients after radical prostatectomy (Prost) and patients diagnosed benign prostatic hyperplasia (BPH) are shown.

Table S2 Confusion table build using results obtained by PLSDA using Ir, Au and Ag Electrodes and 3 LVs. 74 samples were used for calibration and 37 samples for PLSDA model validation. Wilcoxon =0.014 and 0.001 for Cross-Validated.

	validation set	
	PCa	**Control**
Predicted as PCa	**19**	**4**

Predicted as control	**0**	**14**
	19	**18**

Model E) Ir, Rh, Pt, Au and Ag Electrodes,

Supervised discriminant analysis was performed using PLSDA and a maximum number of 7 LVs. Selection of the number of LVs was realized as explained above. Again, the Kennard-Stone method at 66% was used for the split of the data set into the calibration and validation subsets. This resulted in a sensitivity of 81% and a specificity of 100%.

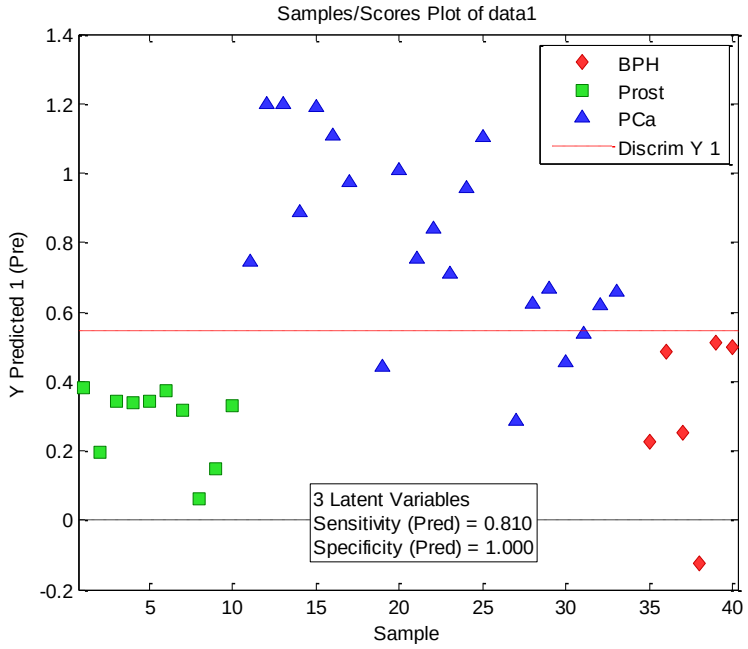

Figure S3. Prediction of validation dataset using 3 LVs using Ir, Rh, Pt, Au and Ag Electrodes. Model has been created using 2 classes (PCa and Control) only for visualization purpose samples from patients after radical prostatectomy (Prost) and patients diagnosed benign prostatic hyperplasia (BPH) are shown.

Table S3. Confusion table build using results obtained by PLSDA using Ir, Rh, Pt, Au and Ag Electrodes and 3 LVs. 74 samples were used for calibration and 37 samples for model validation. Wilcoxon =0.009 and 0.001 for Cross-Validated.

	validation set	
	PCa	**Control**
Predicted as PCa	**17**	**0**
Predicted as control	**4**	**16**
	21	**16**

References

S1. D. J. Balding, M. Bishop and C. Cannings, Handbook of Statistical Genetics, Wiley, New York, 3rd edn., 2007.

S2. R. W. Kennard & L. A. Stone (1969): Computer Aided Design of Experiments, Technometrics, 11:1, 137-148.