**A novel epigenetic signature for early diagnosis in lung cancer**

Dias-Lagares et al.

**Supplementary material and methods**

Study design and participants

Regarding the nature of the controls samples from the minimally and noninvasive cohorts. Bronchial aspirates samples from cancer-free donors (n= 29) were obtained from bronchoscopies that were mainly performed to complete the study of lesions initially suspicious of lung cancer (in 62% of the patients). In 28% of the cases, to assess the diagnosis of lung disorders other than lung cancer and the remaining 10%, as a routine follow-up of patients affected by chronic diseases such as amyloidosis or papyllomatosis. The final diagnosis from patients subjected to study included a wide range of non-malignant lung disorders such as interstitial lung diseases (24%), pulmonary sarcoidosis (20%), lung infection (20%), solitary pulmonary nodule (16%), hemoptisis (8%), hamartomas (8%), and vasculitis (4%). Regarding BALs controls from cancer-free donors (n= 29, only a minority (17% of the patients) were suspicious of lung cancer. BAL samples were obtained principally to assess non-cancerous lung disorders (in 76% of the patients) and the remaining 7% as a routine follow-up. The final diagnosis for these patients included lung infection (41%), interstitial lung diseases (14%), solitary pulmonary nodule (18%), pulmonary sarcoidosis (10%) hemoptisis (7%), foreign body (7%), and thromboembolism (4%). Sputum samples (N= 26) were obtained from cancer-free subjects who participated voluntarily in this study.

**Supplementary Figures and Tables**

**Supplementary figure 1. Flow chart indicating study design**

**Supplementary figure 2. Epigenetic and expression analysis based on histological subtypes from primary tissues of the TCGA database.** (A) DNA methylation values of candidate genes in primary tumor samples subclassified by histological subtypes adenocarcinomas (ADC) and squamous cell carcinomas (SCC). (B) Expression values for the same samples. Non-Tumor in ADC (light grey circle dots), Tumor in ADC (dark grey circle dots), Non-Tumor in SCC (light grey square dots) and Tumor in SCC (dark grey square dots). P-values for all the analyses were calculated using the two-sided Mann–Whitney U test. \*\*\* corresponds to p<0.001.

**Supplementary figure 3. Differentially methylated levels in neighboring CpGs on the selected candidate genes.** Each data point represents the mean β-value of the group (control: blue; adenocarcinoma: red and squamous: green) and whiskers show standard error of the mean (s.e.m). Surrounding CpGs are displayed on X axys (significant and selected CpG is highlighted in green colour). Yellow and brown squares indicated CpG islands and CpG shores regions respectively.

**Supplementary figure 4. Epigenetic signature in lung cancer and non-tumoral paraffins classified based on histological subtypes.** (A) DNA methylation values of candidate genes in paraffins from patients with lung cancer and control donors subclassified by histological subtypes. NT (light grey circle dots) stands for non-tumoral, ADC (grey square dots) for adenocarcinoma and SCC (dark grey triangle dots) for squamous cell carcinomas. P-values for all the analyses were calculated using the two-sided Mann–Whitney U test. \*\*\* corresponds to p<0.001. (B) The area under the curve (AUC) for the each gene in adenocarcinomas. (C) AUC for each gene in squamous cell carcinomas.

**Supplementary figure 5. Results of the epigenetic prediction model for bronchial aspirates (BAS).** (A) Heatmap representing the patients’ methylation profile for the four genes involved in the prediction model. Each column represents a patient and each row one gene. Rows have been ordered according to the results of a hierarchical clustering algorithm and are colored from green to red based on the methylation status (Z-score). Status, stage, cytology results and predicted probability of cancer are depicted in the upper bars with specific color codes. (B) Nomogram for prediction of cancer risk, constructed using the coefficients of our combined logistic regression model as weights. To calculate the probability of cancer, a vertical line straight upward from each factor (*BCAT1, CDO1, TRIM58, ZNF177*) to the Points line has to be drawn. Then, the points from each predictor are summed and with the result, a vertical line is drawn from the Total Points line of the nomogram downwards where the Probability of tumor line is depicted. As a practical example, a patient with the following methylation levels for each gene (*BCAT1*: 2%, *CDO1*: 4%; *TRIM58*: 10% and *ZNF177*: 15%) would get the corresponding points from the Points line: *BCAT11*: 8 points, *CDO1*: 12 points, *TRIM58*: 12 points and *ZNF177*: 30 points. The sum of the four values yielded a total points value of 62. These points correspond to a Probability of Cancer higher than 95%.

Supplementary table 1A. List of cancer-related differentially methylated CpGs comparing adenocarcinomas and non-tumor donors.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TargetID** | **NAME** | **GROUP** | **CpG Island content** | **wilcox.FDR** | **Delta** | **P25P75.diff** |
| cg20399616 | *BCAT1* | Body | Island | 4.8E-02 | 0.48 | 0.20 |
| cg26492446 | *BHLHE23* | TSS200 | Island | 2.8E-02 | 0.40 | 0.18 |
| cg11036833 | *CDO1* | TSS200 | Island | 2.0E-02 | 0.30 | 0.16 |
| cg22399133 | *CRYGD* | 5'UTR | Island | 3.6E-04 | 0.40 | 0.18 |
| cg17276590 | *EMX2OS* | Body | Island | 5.1E-05 | 0.29 | 0.12 |
| cg06947913 | *FAIM2* | TSS200 | Island | 2.9E-02 | 0.40 | 0.23 |
| cg18486102 | *FAIM2* | TSS200 | Island | 1.0E-04 | 0.39 | 0.14 |
| cg16857858 | *HOXA10* | TSS200 | Island | 1.1E-03 | 0.31 | 0.13 |
| cg05311410 | *HOXA11* | Body | Island | 1.8E-02 | 0.27 | 0.12 |
| cg14458834 | *HOXB4* | 1stExon | Island | 3.6E-02 | 0.31 | 0.12 |
| cg08089301 | *HOXB4* | 1stExon | Island | 8.6E-08 | 0.37 | 0.24 |
| cg16422098 | *ITGA8* | TSS200 | Island | 3.0E-07 | 0.29 | 0.11 |
| cg00986824 | *KCNC2* | Body | Island | 6.9E-09 | 0.27 | 0.12 |
| cg04993975 | *LHX1* | TSS200 | Island | 1.6E-03 | 0.32 | 0.13 |
| cg24761507 | *LHX1* | TSS1500 | Island | 3.1E-02 | 0.29 | 0.14 |
| cg10184983 | *SRCIN1* | Body | Island | 2.6E-03 | 0.39 | 0.24 |
| cg13096208 | *ST8SIA3* | 5'UTR | Island | 3.0E-02 | 0.27 | 0.11 |
| cg26673012 | *TBX20* | TSS200 | Island | 4.1E-05 | 0.28 | 0.13 |
| cg25339566 | *TCTEX1D1* | 5'UTR | Island | 9.0E-09 | 0.30 | 0.16 |
| cg05297854 | *TRABD* | TSS1500 | Island | 7.8E-06 | 0.41 | 0.19 |
| cg23054189 | *TRIM58* | 1stExon | Island | 6.8E-05 | 0.40 | 0.27 |
| cg20810478 | *TRIM58* | 1stExon | Island | 1.2E-03 | 0.33 | 0.20 |
| cg07533148 | *TRIM58* | 1stExon | Island | 6.0E-03 | 0.34 | 0.22 |
| cg16021909 | *TRIM58* | Body | Island | 1.0E-02 | 0.29 | 0.11 |
| cg05157140 | *UNCX* | TSS200 | Island | 6.6E-06 | 0.26 | 0.10 |
| cg17138769 | *VAX1* | Body | Island | 5.4E-04 | 0.26 | 0.15 |
| cg08065231 | *ZNF177* | TSS200 | Island | 1.0E-07 | 0.31 | 0.11 |
| cg25397945 | *ZNF529* | TSS200 | Island | 9.4E-03 | 0.39 | 0.22 |

Supplementary table 1B. List of cancer-related differentially methylated CpGs comparing squamous cell carcinomas and non-tumor donors.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TargetID** | **NAME** | **GROUP** | **CpG Island content** | **wilcox.FDR** | **Delta** | **P25P75.diff** |
| cg14996220 | *ALX1* | TSS1500 | Shore | 4.4E-05 | 0.29 | 0.11 |
| cg25026529 | *BARHL2* | TSS1500 | Shore | 5.3E-03 | 0.33 | 0.13 |
| cg20399616 | *BCAT1* | Body | Island | 6.2E-03 | 0.52 | 0.13 |
| cg16998150 | *C1orf114* | TSS200 | Island | 1.0E-03 | 0.32 | 0.13 |
| cg00790098 | *CCDC140* | 5'UTR | Island | 2.5E-05 | 0.34 | 0.18 |
| cg18282849 | *CCDC81* | TSS200 | Island | 5.2E-03 | 0.34 | 0.17 |
| cg11036833 | *CDO1* | TSS200 | Island | 1.2E-04 | 0.38 | 0.10 |
| cg22399133 | *CRYGD* | 5'UTR | Island | 3.5E-05 | 0.61 | 0.38 |
| cg10088985 | *CXCL5* | 5'UTR | Island | 8.5E-05 | 0.36 | 0.21 |
| cg09124223 | *DPP6* | Body | Island | 1.6E-02 | 0.29 | 0.14 |
| cg27277463 | *FAM19A2* | 5'UTR | Island | 1.1E-05 | 0.39 | 0.16 |
| cg00089798 | *FAM19A2* | 5'UTR | Island | 2.4E-02 | 0.29 | 0.10 |
| cg15126544 | *GAD1* | Body | Island | 1.4E-05 | 0.34 | 0.20 |
| cg04912843 | *GIPC2* | 1stExon | Island | 8.9E-03 | 0.32 | 0.14 |
| cg17508991 | *HCK* | TSS1500 | Shore | 2.3E-04 | 0.49 | 0.20 |
| cg18943957 | *HLF* | TSS1500 | Shore | 1.8E-05 | 0.36 | 0.12 |
| cg05311410 | *HOXA11* | TSS1500 | Island | 4.4E-03 | 0.43 | 0.18 |
| cg17466857 | *HOXA11* | TSS1500 | Island | 1.9E-02 | 0.34 | 0.13 |
| cg12600174 | *HOXA9* | TSS200 | Island | 5.5E-05 | 0.33 | 0.16 |
| cg08151857 | *HOXB8* | 1stExon | Island | 2.0E-02 | 0.44 | 0.15 |
| cg10418524 | *HOXD13* | TSS1500 | Island | 2.7E-02 | 0.34 | 0.18 |
| cg18448949 | *HOXD8* | TSS1500 | Island | 2.7E-05 | 0.37 | 0.17 |
| cg25829490 | *HOXD9* | Body | Shore | 1.2E-04 | 0.30 | 0.11 |
| cg09968620 | *HOXD9* | TSS200 | Island | 1.3E-04 | 0.29 | 0.15 |
| cg07881405 | *IRX1* | Body | Island | 2.9E-05 | 0.35 | 0.11 |
| cg08680048 | *ITPKA* | Body | Island | 2.1E-06 | 0.42 | 0.10 |
| cg00986824 | *KCNC2* | Body | Island | 5.4E-06 | 0.35 | 0.16 |
| cg11601252 | *LBXCOR1* | Body | Island | 3.2E-03 | 0.44 | 0.27 |
| cg14093715 | *LHX2* | Body | Island | 3.6E-08 | 0.35 | 0.18 |
| cg05719164 | *LHX4* | Body | Island | 4.1E-08 | 0.39 | 0.10 |
| cg04918402 | *LIMD2* | TSS1500 | Island | 2.4E-03 | 0.31 | 0.16 |
| cg22804236 | *MEI1* | TSS200 | Island | 4.1E-03 | 0.38 | 0.22 |
| cg10296410 | *MEI1* | TSS200 | Shore | 7.3E-04 | 0.28 | 0.14 |
| cg24245418 | *NPY* | Body | Island | 2.1E-05 | 0.33 | 0.13 |
| cg24720571 | *NRN1* | Body | Island | 3.5E-03 | 0.29 | 0.13 |
| cg12162138 | *ODZ4* | Body | Island | 4.7E-06 | 0.35 | 0.21 |
| cg27122213 | *OTX2* | 5'UTR | Island | 9.2E-04 | 0.36 | 0.20 |
| cg10042106 | *OTX2* | 5'UTR | Island | 2.0E-03 | 0.32 | 0.17 |
| cg05477457 | *PALLD* | Body | Island | 3.7E-03 | 0.46 | 0.23 |
| cg13482308 | *PAX5* | Body | Island | 1.3E-03 | 0.38 | 0.21 |
| cg16616521 | *PAX6* | Body | Island | 5.4E-06 | 0.31 | 0.13 |
| cg18617005 | *PCDHGs* | Body | Island | 1.5E-04 | 0.37 | 0.14 |
| cg18507379 | *PCDHGs* | Body | Island | 3.5E-03 | 0.37 | 0.16 |
| cg23663774 | *PITX2* | 5'UTR | Island | 4.9E-03 | 0.45 | 0.28 |
| cg02725370 | *PITX2* | Body | Shore | 1.2E-02 | 0.30 | 0.14 |
| cg13645732 | *PPP1R13L* | 5'UTR | Island | 3.6E-04 | 0.38 | 0.17 |
| cg12993163 | *SHOX2* | Body | Island | 6.5E-05 | 0.42 | 0.23 |
| cg08074534 | *SIM1* | Body | Island | 1.1E-08 | 0.28 | 0.11 |
| cg19456540 | *SIX6* | 1stExon | Island | 1.4E-03 | 0.36 | 0.14 |
| cg10439765 | *SLC12A5* | 1stExon | Island | 1.1E-03 | 0.27 | 0.15 |
| cg26107890 | *SLC12A8* | Body | Island | 1.6E-03 | 0.58 | 0.14 |
| cg12788878 | *SLC12A8* | Body | Island | 7.8E-04 | 0.32 | 0.13 |
| cg24928391 | *SOX17* | TSS200 | Island | 4.0E-08 | 0.31 | 0.10 |
| cg17964510 | *SP9* | TSS200 | Island | 7.8E-04 | 0.39 | 0.19 |
| cg13096208 | *ST8SIA3* | 5'UTR | Island | 2.0E-03 | 0.37 | 0.16 |
| cg25528916 | *STAP2* | Body | Island | 6.7E-03 | 0.47 | 0.17 |
| cg05255522 | *TFAP2D* | Body | Island | 1.9E-05 | 0.26 | 0.10 |
| cg22512438 | *TRH* | 5'UTR | Island | 2.3E-04 | 0.31 | 0.10 |
| cg23054189 | *TRIM58* | 1stExon | Island | 5.1E-04 | 0.48 | 0.35 |
| cg20146541 | *TRIM58* | 1stExon | Island | 4.2E-02 | 0.48 | 0.36 |
| cg07533148 | *TRIM58* | 1stExon | Island | 6.6E-07 | 0.44 | 0.28 |
| cg20810478 | *TRIM58* | 1stExon | Island | 1.9E-04 | 0.42 | 0.26 |
| cg16021909 | *TRIM58* | Body | Island | 2.3E-04 | 0.41 | 0.17 |
| cg17138769 | *VAX1* | Body | Island | 2.1E-05 | 0.42 | 0.24 |
| cg22344703 | *WNT6* | Body | Island | 1.1E-06 | 0.37 | 0.19 |
| cg06862374 | *WNT6* | Body | Island | 2.4E-04 | 0.30 | 0.12 |
| cg05661282 | *ZNF154* | 5'UTR | Island | 2.0E-05 | 0.49 | 0.30 |
| cg03142586 | *ZNF154* | Body | Shore | 1.5E-02 | 0.34 | 0.11 |
| cg08065231 | *ZNF177* | TSS200 | Island | 1.5E-06 | 0.45 | 0.24 |
| cg04332534 | *ZNF382* | 5'UTR | Island | 1.3E-07 | 0.34 | 0.19 |
| cg05020604 | *ZNF529* | TSS200 | Island | 2.8E-02 | 0.39 | 0.24 |
| cg23035602 | *ZNF761* | TSS200 | Island | 2.4E-03 | 0.29 | 0.17 |
| cg14587524 | *ZNF781* | TSS200 | Island | 3.9E-07 | 0.45 | 0.28 |

Supplementary table 1C. List of cancer-related differentially methylated CpGs comparing lung cancer (ADC+SCC) and non-tumor donors.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **TargetID** | **NAME** | **GROUP** | **CpG Island content** | **wilcox.FDR** | **Delta** | **P25P75.diff** |
| cg20399616 | *BCAT1* | Body | Island | 2.2E-02 | 0.49 | 0.18 |
| cg26492446 | *BHLHE23* | TSS200 | Island | 3.5E-02 | 0.37 | 0.11 |
| cg11036833 | *CDO1* | TSS200 | Island | 4.3E-03 | 0.32 | 0.16 |
| cg22399133 | *CRYGD* | 5'UTR | Island | 9.3E-05 | 0.44 | 0.19 |
| cg27277463 | *FAM19A2* | 5'UTR | Island | 9.2E-03 | 0.34 | 0.14 |
| cg05311410 | *HOXA11* | TSS1500 | Island | 8.7E-03 | 0.30 | 0.13 |
| cg12600174 | *HOXA9* | TSS200 | Island | 3.3E-06 | 0.26 | 0.12 |
| cg08089301 | *HOXB4* | 1stExon | Island | 1.8E-08 | 0.37 | 0.22 |
| cg00986824 | *KCNC2* | Body | Island | 9.4E-09 | 0.28 | 0.13 |
| cg27122213 | *OTX2* | 5'UTR | Island | 1.1E-03 | 0.26 | 0.13 |
| cg19456540 | *SIX6* | 1stExon | Island | 5.2E-04 | 0.30 | 0.11 |
| cg10184983 | *SRCIN1* | Body | Island | 7.6E-04 | 0.37 | 0.14 |
| cg13096208 | *ST8SIA3* | 5'UTR | Island | 1.1E-02 | 0.29 | 0.12 |
| cg26673012 | *TBX20* | TSS200 | Island | 1.9E-05 | 0.28 | 0.13 |
| cg25339566 | *TCTEX1D1* | 5'UTR | Island | 2.9E-08 | 0.30 | 0.14 |
| cg05297854 | *TRABD* | TSS1500 | Island | 5.0E-06 | 0.40 | 0.11 |
| cg23054189 | *TRIM58* | 1stExon | Island | 5.2E-05 | 0.42 | 0.28 |
| cg07533148 | *TRIM58* | 1stExon | Island | 4.4E-04 | 0.36 | 0.23 |
| cg20810478 | *TRIM58* | 1stExon | Island | 3.9E-04 | 0.35 | 0.21 |
| cg16021909 | *TRIM58* | Body | Island | 2.6E-03 | 0.32 | 0.12 |
| cg17138769 | *VAX1* | Body | Island | 1.2E-04 | 0.29 | 0.16 |
| cg08065231 | *ZNF177* | TSS200 | Island | 5.2E-08 | 0.33 | 0.12 |
| cg25397945 | *ZNF529* | TSS200 | Island | 1.3E-02 | 0.40 | 0.27 |