**SUPPLEMENTARY MATERIAL**

**Figure Legends**

**Supplementary Figure 1:** Distributions of the 13 selected methylated DNA marker candidates across case and control subsets from tissue validation set and cyst fluid pilot.

**Supplementary Figure 2:** Relative ranking of CF markers using random forest regression. The decrease in accuracy an MDM causes is determined by randomly permuting group classifications along with randomly selecting MDM entry into an individual tree at each iteration of the random forest algorithm. The accuracy of an MDM will have larger decreases in the Gini index when excluded from tree or when permutation group classifications. The changes in the Gini index for each variable are aggregated across all iterations and normalized at the end to a relative ranking scale.

**Supplementary Figure 3:** Pilot trained rPart and rForest ROC curves for MDMs in the cyst fluid validation cohort

**Supplementary Table 1:** Discrimination of top candidate methylated DNA markers in tissue discovery set: cases with HGD/cancer vs controls with no dysplasia/LGD. High-individual AUCs, high fold change (FC) and significant p-values were considered for selection of markers to be carried forward for further exploration in independent tissue validation and subsequent cyst fluid pilot testing.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Methylated DNA Marker** |  | **AUC** | **FC** | **p-value** |
| CLEC11A |  | 0.96 | 381.88 | 5.79E-08 |
| GRIN2D |  | 0.94 | 5.90 | 4.66E-08 |
| VWC2 |  | 0.94 | 611.99 | 2.21E-07 |
| ST8SIA1 |  | 0.9 | 77.18 | 0.048 |
| c13orf18 |  | 0.87 | 5.02 | 0.000521007 |
| AK055957 | 0.87 | 568.49 | 1.77E-05 |
| FER1L4 |  | 0.85 | 8.54 | 8.03E-05 |
| ELMO1 |  | 0.82 | 7611.88 | 0.008461 |
| TBX15 |  | 0.81 | 35.03 | 6.53E-05 |
| CD1D |  | 0.81 | >10000 | 0.023673984 |
| EMX1 |  | 0.78 | 353.24 | 0.00240684 |
| SP9 |  | 0.78 | 62.49 | 0.005096101 |
| BMP3 |  | 0.78 | 961.95 | 0.005443797 |
| NDRG4 |  | 0.78 | >10000 | 0.019846862 |
| HOXA1 |  | 0.77 | 17.69 | 0.000206205 |
| DLX4 |  | 0.77 | 13.51 | 0.003187683 |
| PRKCB |  | 0.75 | 166.30 | 0.08341 |
| ABCB1 |  | 0.73 | 29.55 | 0.01351915 |
| ZNF781 |  | 0.66 | 1121.48 | 0.00831417 |
| C4orf31\* |  | 0.87 | 6.244 | 0.0001257 |
| FRMD4A\* |  | 0.86 | 16.99 | 0.004081 |
| ST6GAL2\* |  | 0.83 | 13.4 | 0.00000119 |
| CXCR4\* |  | 0.81 | 10.17 | 0.00001124 |
| KCNQ5\* |  | 0.75 | 5.254 | 0.04692 |
| ZNF682\* |  | 0.70 | 6.038 | 0.0006942 |

\*Markers derived from HGD/LGD comparisions

**Supplementary Table 2:** Discrimination of case from control tissues by top candidate methylated DNA markers in biological tissue validation in an independent sample set. High-individual AUCs justified further exploration in subsequent cyst fluid pilot testing.

|  |  |
| --- | --- |
| **Methylated DNA marker** | **AUC (95% CI)** |
| *ACBC1* | 0.86 (0.79-0.92) |
| *BMP3* | 0.83 (0.76-0.91) |
| *CD1D* | 0.82 (0.74-0.89) |
| *CLEC11A* | 0.88 (0.82-0.94) |
| *DLX4* | 0.86 (0.80-0.92) |
| *ELMO1* | 0.87 (0.8-0.94) |
| *EMX1* | 0.87 (0.81-0.93) |
| *NDRG4* | 0.75 (0.66-0.84) |
| *PRKCB* | 0.88 (0.82-0.94) |
| *SP9* | 0.85 (0.78-0.92) |
| *ST8SIA1* | 0.86 (0.79-0.92) |
| *TBX15* | 0.88 (0.82-0.94) |
| *VWC2* | 0.88 (0.82-0.94) |
| *C13orf18* | 0.81 (0.73-0.88) |
| *C4orf31* | 0.68 (0.57-0.78) |
| *CHR12.133* | 0.80 (0.72-0.89) |
| *CXCR4* | 0.57 (0.47-0.68) |
| *FER1L4* | 0.83 (0.76-0.9) |
| *FRMD4A* | 0.84 (0.76-0.91) |
| *GRIN2D* | 0.84 (0.77-0.91) |
| *HOXA1* | 0.81 (0.74-0.89) |
| *KCNQ5* | 0.74 (0.65-0.82) |
| *ST6GAL2* | 0.69 (0.59-0.79) |
| *ZNF682* | 0.69 (0.59-0.8) |
| *ZNF781* | 0.80 (0.71-0.88) |

**Supplementary Table 3:**  Discrimination of selected methylated DNA markers for detection of HGD/cancer vs no dysplasia/LGD in pancreatic cysts: AUCs in tissue validation set and cyst fluid pilot and validation.

|  |  |
| --- | --- |
| **Methylated DNA marker** | **AUCs (95% CI)** |
|  | Tissue Validation | Cyst Fluid Pilot | Cyst fluid validation\* |
| *BMP3* | 0.83 (0.76-0.91) | 0.92 (0.85-0.98) | 0.92 (0.85-0.99) |
| *TBX15* | 0.88 (0.82-0.94) | 0.92 (0.85-0.99) | 0.9 (0.82-0.98) |
| *CLEC11A* | 0.88 (0.82-0.94) | 0.91 (0.85-0.97) | 0.9 (0.82-0.98) |
| *ST8SIA1* | 0.86 (0.79-0.92) | 0.90 (0.84-0.96) | 0.89 (0.81-0.98) |
| *DLX4* | 0.86 (0.80-0.92) | 0.87 (0.80-0.94) | 0.85 (0.75-0.94) |
| *PRKCB* | 0.88 (0.82-0.94) | 0.87 (0.79-0.95) | 0.87 (0.78-0.95) |
| *ELMO1* | 0.87 (0.80-0.94) | 0.86 (0.78-0.94) | 0.89 (0.82-0.96) |
| *EMX1* | 0.87 (0.81-0.93) | 0.86 (0.76-0.96) | 0.9 (0.84-0.97) |
| *VWC2* | 0.88 (0.82-0.94) | 0.84 (0.73-0.95) | 0.9 (0.82-0.97) |
| *CD1D* | 0.82 (0.74-0.89) | 0.84 (0.71-0.96) | 0.91 (0.84-0.98) |
| *SP9* | 0.85 (0.78-0.92) | 0.82 (0.71-0.93) | 0.86 (0.78-0.95) |
| *NDRG4* | 0.75 (0.66-0.84) | 0.80 (0.69-0.92) | 0.81 (0.72-0.9) |
| *ACBC1* | 0.86 (0.79-0.92) | 0.71 (0.56-0.86) | 0.86 (0.78-0.94) |

**\* p value >0.05 for each individual MDM AUCs when compared to pilot phase results**

**Supplementary Table 4:**  Effect of age and sex on discrimination by selected methylated DNA markers (MDMs) in cyst fluid pilot: AUCs (95% CI) shown.

|  |  |  |  |
| --- | --- | --- | --- |
| MDM | Age<60 | Age≥60 | P value |
| ndrg4 | 0.74 (0.52-0.96) | 0.82 (0.69-0.95) | 0.5386 |
| bmp3 | 0.98 (0.94-1.00) | 0.87 (0.76-0.98) | 0.0726 |
| sp9 | 0.75 (0.53-0.98) | 0.83 (0.70-0.97) | 0.5319 |
| dlx4 | 0.92 (0.84-1.00) | 0.82 (0.70-0.94) | 0.1477 |
| abcb1 | 0.50 (0.17-0.83) | 0.77 (0.63-0.92) | 0.1385 |
| cd1d | 0.85 (0.57-1.00) | 0.82 (0.66-0.97) | 0.8221 |
| clec11a | 0.98 (0.96-1.00) | 0.86 (0.75-0.96) | 0.0212 |
| emx1 | 0.97 (0.92-1.00) | 0.80 (0.66-0.95) | 0.0356 |
| prkcb | 0.91 (0.83-1.00) | 0.85 (0.74-0.96) | 0.4047 |
| st8sia1 | 0.93 (0.85-1.00) | 0.90 (0.81-0.98) | 0.5919 |
| vwc2 | 0.95 (0.87-1.00) | 0.78 (0.61-0.94) | 0.0692 |
| tbx15 | 0.95 (0.90-1.00) | 0.90 (0.81-1.00) | 0.3594 |
| elmo1 | 0.88 (0.76-1.00) | 0.84 (0.72-0.95) | 0.6130 |

|  |  |  |  |
| --- | --- | --- | --- |
| MDM | Male | Female | P value |
| ndrg4 | 0.85 (0.71-0.98) | 0.72 (0.54-0.89) | 0.2560 |
| bmp3 | 0.87 (0.76-0.98) | 0.93 (0.84-1.00) | 0.4090 |
| sp9 | 0.76 (0.59-0.94) | 0.84 (0.67-1.00) | 0.5428 |
| dlx4 | 0.78 (0.64-0.93) | 0.91 (0.82-1.00) | 0.1419 |
| abcb1 | 0.69 (0.5-0.87) | 0.67 (0.39-0.95) | 0.9217 |
| cd1d | 0.81 (0.65-0.97) | 0.81 (0.58-1.00) | 0.9910 |
| clec11a | 0.85 (0.73-0.96) | 0.93 (0.84-1.00) | 0.2563 |
| emx1 | 0.76 (0.6-0.93) | 0.91 (0.81-1.00) | 0.1385 |
| prkcb | 0.79 (0.64-0.93) | 0.88 (0.78-0.98) | 0.2809 |
| st8sia1 | 0.85 (0.73-0.97) | 0.93 (0.85-1.00) | 0.3042 |
| vwc2 | 0.75 (0.58-0.92) | 0.88 (0.71-1.00) | 0.3053 |
| tbx15 | 0.91 (0.81-1.00) | 0.90 (0.77-1.00) | 0.9757 |
| elmo1 | 0.78 (0.64-0.92) | 0.86 (0.72-1.00) | 0.4085 |

**Supplementary Table 5:** Gene location, name, and function for methylated DNA candidates that appear to be discriminant for HGD/cancer in cyst fluid.

|  |  |  |  |
| --- | --- | --- | --- |
| **Methylated DNA marker** | Chromosomal location | Name | Function |
| ABCB1 | 7 |  ATP-binding cassette, sub-family B | transport |
| BMP3 | 4 | Bone morphogenetic protein 3 | growth factor |
| CD1D | 1 | Antigen-presenting glycoprotein CD1d | immune response |
| CLEC11 | 19 | C-type lectin domain family 11, member A | growth factor |
| DLX4 | 17 | Distal-less homeobox 4 | transcriptional regulation |
| ELMO1 | 7 | Engulfment and cell motility 1 | chemokine signaling |
| EMX1 | 2 | Empty spiracles homeobox 1 | transcriptional regulation |
| NDRG4 | 16 | N-myc downstream regulated gene 4 | cell cycle/signaling |
| PRKCB | 16 | Protein kinase C, beta | signal transduction |
| SP9 | 2 | Sp9 transcription factor | transcriptional regulation |
| ST8SIA1 | 12 | ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 1 | carbohydrate metabolism |
| TBX15 | 1 | T-box 15 | transcriptional regulation |
| VWC2 | 7 | von Willebrand factor C domain containing 2 | signal transduction |