**Table 1 Characteristics of the CRC patients included in this study**

| **Characteristics** | **Patient Distribution** | **Patient Distribution** |
| --- | --- | --- |
| **Cohort 1 (N = 104)** | **Cohort 2 (N = 114)** |
| **Age** | 66 (IQR = 62 to 74) | 68 (IQR = 60 to 75) |
| **Gender** |  |  |
| Male | 71 | 75 |
| Female | 33 | 39 |
| **Subtype a** |  |  |
| Colon | 55 | 68 |
| Rectal | 49 | 45 |
| **UICC Stage b** |  |  |
| I | 18 | 28 |
| II | 35 | 38 |
| III | 40 | 38 |
| IV | 11 | 9 |
| **Tumor invasion depthc** |  |  |
| T1 | 5 | 7 |
| T2 | 21 | 22 |
| T3 | 70 | 69 |
| T4 | 6 | 15 |
| Tx | 2 | 1 |
| **Lymph node involvement c** |  |  |
| N0 | 57 | 72 |
| N1 | 29 | 25 |
| N2 | 11 | 15 |
| N3 | 5 | 2 |
| Nx | 2 |  |
| **Distant metastasis c** |  |  |
| M0 | 93 | 105 |
| M1 | 11 | 9 |
| **KRAS mutation d** |  |  |
| Positive | 50 | 52 |
| Negative | 52 | 62 |

cTNM Stages were assessed by the seventh edition of the TNM classification criteria.

**Table 2 The mean methylation status of the 5 genomic regions in replication cohort 1**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Genomic**  **Regiona** | **Gene** | **McaMc** | **McoMc** | **P valued** | **FDR** | **OR** | **95% CIe** | **Sensf** | **Specf** | **AUCf** |
| 6:152129591-152129791 | *ESR1* | 0.36 | 0.07 | 1.00 × 10-23 | 6.10 × 10-23 | 7.15 | 5.04 - 9.83 | 0.87 | 0.88 | 0.93 |
| 19:58951599-58951728 | *ZNF132* | 0.36 | 0.04 | 1.90 × 10-22 | 5.60 × 10-22 | 9.13 | 6.10 - 13.10 | 0.83 | 0.98 | 0.91 |
| 19:44952604-44952808 | *ZNF229* | 0.32 | 0.11 | 6.30 × 10-17 | 6.30 × 10-17 | 6.34 | 4.57 - 8.49 | 0.69 | 0.96 | 0.85 |
| 19:56879613-56879735 | *ZNF542* | 0.43 | 0.14 | 1.10 × 10-20 | 1.60 × 10-20 | 5.50 | 4.03 - 7.30 | 0.79 | 0.94 | 0.90 |
| 19:53758048-53758164 | *ZNF677* | 0.51 | 0.25 | 5.90 × 10-18 | 7.10 × 10-18 | 4.51 | 3.31 - 5.94 | 0.69 | 0.94 | 0.87 |
| 17:75369456-75369630 | *SEPT-9* | 0.46 | 0.10 | 1.10 × 10-21 | 2.20 × 10-21 | 4.46 | 3.30 - 5.86 | 0.86 | 0.87 | 0.91 |

aGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here is based on the hg19 version of the genome. bNo.CpG sites represents the number of the CpG sites in each region. cMcaM represents the mean methylation percentage of the cases in each region, which consisting of several CpG sites, while the McoM represents the mean methylation percentage of the controls in each region. dP value is calculated through the wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. e OR and 95% CI were conducted through logistic regression. eSens = sensitivity, while Spec = specificity, AUC = area under curve. The sensitivity, specificity as well as the AUC were calculated through a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status.

**Table 3 The mean methylation status of the 5 genomic regions in the KRAS+ and KRAS- samples of replication cohort 1**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Genomic**  **Regiona** | **Gene** | **McaMb** | **McoMc** | **P valued** | **FDR** | **OR** | **95% CIe** | **Sensf** | **Specf** | **AUCf** |
| **KRAS+** | 6:152129591-152129791 | *ESR1* | 0.46 | 0.06 | 1.70 × 10-14 | 5.00 × 10-14 | 21.8 | 11.00 - 43.50 | 0.98 | 1.00 | 1.00 |
| 19:58951599-58951728 | *ZNF132* | 0.48 | 0.04 | 2.80 × 10-14 | 5.00 × 10-14 | 28.7 | 13.20 - 28.70 | 0.98 | 1.00 | 0.99 |
| 19:44952604-44952808 | *ZNF229* | 0.41 | 0.11 | 1.30 × 10-13 | 1.90 × 10-13 | 14.4 | 8.59 - 23.70 | 0.98 | 0.89 | 0.98 |
| 19:56879613-56879735 | *ZNF542* | 0.55 | 0.13 | 1.30 × 10-14 | 5.00 × 10-14 | 62.6 | 20.80 - 62.60 | 0.98 | 1.00 | 1.00 |
| 19:53758048-53758164 | *ZNF677* | 0.59 | 0.26 | 2.10 × 10-13 | 2.50 × 10-13 | 12.3 | 7.12 - 20.40 | 0.86 | 0.97 | 0.98 |
| 17:75369456-75369630 | *SEPT-9* | 0.55 | 0.08 | 3.70 × 10-13 | 3.70 × 10-13 | 7.86 | 4.74 - 12.90 | 0.91 | 0.97 | 0.97 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **KRAS-** | 6:152129591-152129791 | *ESR1* | 0.28 | 0.07 | 1.10 × 10-9 | 6.60 × 10-9 | 6.18 | 3.48 - 9.83 | 0.76 | 0.87 | 0.86 |
| 19:58951599-58951728 | *ZNF132* | 0.25 | 0.04 | 4.60 × 10-9 | 1.40 × 10-8 | 7.32 | 3.96 - 12.30 | 0.70 | 0.98 | 0.85 |
| 19:44952604-44952808 | *ZNF229* | 0.25 | 0.11 | 4.90 × 10-5 | 4.90 × 10-5 | 4.37 | 2.39 - 6.88 | 0.52 | 0.96 | 0.74 |
| 19:56879613-56879735 | *ZNF542* | 0.33 | 0.13 | 1.90 × 10-7 | 2.90 × 10-7 | 4.31 | 2.56 - 6.49 | 0.56 | 0.98 | 0.81 |
| 19:53758048-53758164 | *ZNF677* | 0.44 | 0.24 | 1.60 × 10-6 | 1.90 × 10-6 | 3.30 | 1.93 - 5.00 | 0.72 | 0.82 | 0.79 |
| 17:75369456-75369630 | *SEPT-9* | 0.39 | 0.11 | 7.50 × 10-9 | 1.50 × 10-8 | 3.77 | 2.34 - 5.56 | 0.82 | 0.80 | 0.84 |

aGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here is based on the hg19 version of the genome. bMcaM represents the mean methylation percentage of the cases in each region, which consisting of several CpG sites, while the McoMc represents the mean methylation percentage of the controls in each region. dP value is calculated using wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple test correction. eOR and 95% CI were conducted through logistic regression. eSens = sensitivity, while Spec = specificity, AUC = area under curve. The sensitivity, specificity as well as the AUC were calculated through a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status.

**Table 4 The mean methylation status of the 5 genomic regions in replication cohort 2**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Genomic**  **Regiona** | **Gene** | **McaMc** | **McoMc** | **P valued** | **FDR** | **OR** | **95% CIe** | **Sensf** | **Specf** | **AUCf** |
| 6:152129591-152129791 | *ESR1* | 0.41 | 0.071 | 8.50 × 10-24 | 2.10 × 10-23 | 5.37 | 3.91 - 7.18 | 0.79 | 0.95 | 0.90 |
| 19:58951599-58951728 | *ZNF132* | 0.42 | 0.036 | 3.60 × 10-28 | 1.80 × 10-27 | 8.28 | 5.67 - 11.70 | 0.83 | 0.97 | 0.94 |
| 19:44952604-44952808 | *ZNF229* | 0.42 | 0.116 | 6.00 × 10-18 | 6.00 × 10-18 | 4.12 | 2.95 - 5.56 | 0.71 | 0.93 | 0.85 |
| 19:56879613-56879735 | *ZNF542* | 0.46 | 0.139 | 7.20 × 10-19 | 1.20 × 10-18 | 3.71 | 2.72 - 4.90 | 0.78 | 0.89 | 0.86 |
| 19:53758048-53758164 | *ZNF677* | 0.57 | 0.27 | 1.40 × 10-18 | 1.80 × 10-18 | 3.42 | 2.54 - 4.46 | 0.78 | 0.86 | 0.85 |
| 17:75369456-75369630 | *SEPT-9* | 0.52 | 0.096 | 6.50 × 10-25 | 6.50 × 10-25 | 3.66 | 2.73 - 4.80 | 0.77 | 0.94 | 0.91 |

aGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here is based on the hg19 version of the genome. bNo.CpG sites represents the number of the CpG sites in each region. cMcaM represents the mean methylation percentage of the cases in each region, which consisting of several CpG sites, while the McoM represents the mean methylation percentage of the controls in each region. dP value is calculated through the wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. e OR and 95% CI were conducted through logistic regression. eSens = sensitivity, while Spec = specificity, AUC = area under curve. The sensitivity, specificity as well as the AUC were calculated through a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status.

**Table 5 The mean methylation status of the 5 genomic regions in the KRAS+ and KRAS- samples of replication cohort 2**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Genomic**  **Regiona** | **Gene** | **McaMc** | **McoMc** | **P valued** | **FDR** | **OR** | **95% CIe** | **Sensf** | **Specf** | **AUCf** |
| **KRAS+** | 6:152129591-152129791 | *ESR1* | 0.52 | 0.06 | 8.60 × 10-18 | 5.10 × 10-17 | 8.34 | 5.18 - 12.9 | 0.92 | 0.98 | 0.95 |
| 19:58951599-58951728 | *ZNF132* | 0.54 | 0.03 | 2.00 × 10-16 | 6.10 × 10-16 | 30.70 | 15.40 - 66.70 | 0.98 | 0.98 | 1.00 |
| 19:44952604-44952808 | *ZNF229* | 0.52 | 0.11 | 2.40 × 10-12 | 2.40 × 10-12 | 7.79 | 4.57 - 12.20 | 0.90 | 0.91 | 0.92 |
| 19:56879613-56879735 | *ZNF542* | 0.59 | 0.12 | 5.80 × 10-16 | 1.20 × 10-15 | 11.10 | 6.85 - 17.70 | 1.00 | 0.93 | 0.99 |
| 19:53758048-53758164 | *ZNF677* | 0.67 | 0.24 | 3.60 × 10-15 | 4.40 × 10-15 | 10.10 | 6.25 - 15.90 | 0.94 | 0.93 | 0.98 |
| 17:75369456-75369630 | *SEPT-9* | 0.61 | 0.08 | 1.80 × 10-15 | 2.70 × 10-15 | 7.89 | 4.87 - 12.70 | 0.90 | 0.98 | 0.98 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **KRAS-** | 6:152129591-152129791 | *ESR1* | 0.33 | 0.08 | 3.50 × 10-10 | 7.40 × 10-10 | 3.90 | 2.37 - 5.94 | 0.67 | 0.95 | 0.85 |
| 19:58951599-58951728 | *ZNF132* | 0.31 | 0.04 | 5.70 × 10-12 | 3.40 × 10-11 | 5.37 | 3.09 - 8.61 | 0.85 | 0.84 | 0.88 |
| 19:44952604-44952808 | *ZNF229* | 0.32 | 0.13 | 2.50 × 10-6 | 3.70 × 10-6 | 2.78 | 1.58 - 4.30 | 0.69 | 0.80 | 0.76 |
| 19:56879613-56879735 | *ZNF542* | 0.34 | 0.16 | 1.10 × 10-4 | 1.10 × 10-4 | 2.02 | 1.07 - 3.20 | 0.58 | 0.85 | 0.71 |
| 19:53758048-53758164 | *ZNF677* | 0.49 | 0.30 | 7.30 × 10-5 | 8.80 × 10-5 | 1.98 | 1.08 - 3.02 | 0.49 | 0.96 | 0.72 |
| 17:75369456-75369630 | *SEPT-9* | 0.43 | 0.11 | 3.70 × 10-10 | 7.40 × 10-10 | 2.43 | 1.51 - 3.59 | 0.78 | 0.78 | 0.85 |

aGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here is based on the hg19 version of the genome. bMcaM represents the mean methylation percentage of the cases in each region, which consisting of several CpG sites, while the McoMc represents the mean methylation percentage of the controls in each region. dP value is calculated using wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple test correction. eOR and 95% CI were conducted through logistic regression. eSens = sensitivity, while Spec = specificity, AUC = area under curve. The sensitivity, specificity as well as the AUC were calculated through a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status.

**Table 6 The mean methylation status of the 5 genomic regions in the KRAS+ and KRAS- samples of replication cohort 1 and cohort 2**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Genomic**  **Regiona** | **Gene** | **McaMc** | **McoMc** | **P valued** | **FDR** | **OR** | **95% CIe** | **Sensf** | **Specf** | **AUCf** |
| **KRAS+** | 6:152129591-152129791 | *ESR1* | 0.49 | 0.06 | 8.00 × 10-27 | 9.60 × 10-27 | 11.00 | 7.53 - 15.60 | 0.96 | 0.98 | 0.97 |
| 19:58951599-58951728 | *ZNF132* | 0.51 | 0.03 | 2.10 × 10-29 | 1.20 × 10-28 | 29.00 | 17.6 - 47.90 | 0.98 | 0.99 | 1.00 |
| 19:44952604-44952808 | *ZNF229* | 0.47 | 0.11 | 1.70 × 10-24 | 1.70 × 10-24 | 10.20 | 7.19 - 14.00 | 0.90 | 0.93 | 0.95 |
| 19:56879613-56879735 | *ZNF542* | 0.57 | 0.13 | 4.00 × 10-29 | 1.20 × 10-28 | 14.30 | 9.60 - 21.50 | 0.99 | 0.95 | 0.99 |
| 19:53758048-53758164 | *ZNF677* | 0.63 | 0.25 | 3.40 × 10-27 | 5.10 × 10-27 | 10.90 | 7.64 - 15.30 | 0.90 | 0.95 | 0.98 |
|  | 17:75369456-75369630 | *SEPT-9* | 0.58 | 0.08 | 2.40 × 10-27 | 4.70 × 10-27 | 7.85 | 5.55 - 11.00 | 0.91 | 0.96 | 0.98 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **KRAS-** | 6:152129591-152129791 | *ESR1* | 0.31 | 0.08 | 2.60 × 10-18 | 7.90 × 10-18 | 4.63 | 3.21 - 6.39 | 0.75 | 0.83 | 0.85 |
| 19:58951599-58951728 | *ZNF132* | 0.28 | 0.04 | 1.30 × 10-19 | 7.70 × 10-19 | 6.10 | 4.09 - 8.73 | 0.70 | 0.97 | 0.87 |
| 19:44952604-44952808 | *ZNF229* | 0.29 | 0.12 | 6.90 × 10-10 | 8.20 × 10-10 | 3.30 | 2.20 - 4.60 | 0.54 | 0.94 | 0.75 |
| 19:56879613-56879735 | *ZNF542* | 0.33 | 0.15 | 1.50 × 10-10 | 2.30 × 10-10 | 2.78 | 1.87 - 3.84 | 0.50 | 0.96 | 0.76 |
| 19:53758048-53758164 | *ZNF677* | 0.46 | 0.27 | 1.80 × 10-09 | 1.80 × 10-09 | 2.38 | 1.62 - 3.25 | 0.71 | 0.73 | 0.74 |
|  | 17:75369456-75369630 | *SEPT-9* | 0.41 | 0.11 | 2.30 × 10-17 | 4.70 × 10-17 | 2.94 | 2.11 - 3.92 | 0.77 | 0.80 | 0.84 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **Total** | 6:152129591-152129791 | *ESR1* | 0.39 | 0.07 | 1.20 × 10-45 | 3.50 × 10-45 | 6.06 | 4.81 - 7.52 | 0.78 | 0.95 | 0.91 |
| 19:58951599-58951728 | *ZNF132* | 0.39 | 0.04 | 6.30 × 10-49 | 3.80 × 10-48 | 8.67 | 6.60 - 11.20 | 0.83 | 0.97 | 0.93 |
| 19:44952604-44952808 | *ZNF229* | 0.37 | 0.11 | 3.50 × 10-33 | 3.50 × 10-33 | 5.04 | 3.99 - 6.24 | 0.70 | 0.94 | 0.85 |
| 19:56879613-56879735 | *ZNF542* | 0.44 | 0.14 | 1.00 × 10-37 | 1.50 × 10-37 | 4.42 | 3.56 - 5.39 | 0.79 | 0.90 | 0.87 |
| 19:53758048-53758164 | *ZNF677* | 0.54 | 0.26 | 1.60 × 10-34 | 1.90 × 10-34 | 3.85 | 3.11 - 4.67 | 0.73 | 0.90 | 0.86 |
| 17:75369456-75369630 | *SEPT-9* | 0.49 | 0.10 | 5.50 × 10-45 | 1.10 × 10-45 | 4.02 | 3.27 - 4.02 | 0.83 | 0.87 | 0.91 |

aGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here is based on the hg19 version of the genome. bNo.CpG sites represents the number of the CpG sites in each region. cMcaM represents the mean methylation percentage of the cases in each region, which consisting of several CpG sites, while the McoM represents the mean methylation percentage of the controls in each region. dP value is calculated through the wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. e OR and 95% CI were conducted through logistic regression. eSens = sensitivity, while Spec = specificity, AUC = area under curve. The sensitivity, specificity as well as the AUC were calculated through a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status.