Figure S9

**a**

KIRC 138 samples set A → Random split into 2 sets A1, A2 → set A1

```
smRNAs exon 1
Differentially expressed
FDR<0.05

735 features
```

**GO Term - Biological Process** | **P value** | **Benjamini**
--- | --- | ---
Immunity and defense | 5.80E-06 | 9.10E-04
T-cell mediated immunity | 1.70E-04 | 1.30E-02
MHCI-mediated immunity | 3.90E-03 | 1.80E-01
Oncogenesis | 5.70E-03 | 2.00E-01
Blood clotting | 6.50E-03 | 1.90E-01
MHCI-mediated immunity | 1.30E-02 | 2.90E-01
Cell proliferation and differentiation | 2.50E-02 | 4.40E-01
Homeostasis | 2.90E-02 | 4.30E-01
Oncogene | 3.90E-02 | 5.00E-01
Hematopoiesis | 4.30E-02 | 5.00E-01
Developmental processes | 5.70E-02 | 5.70E-01
Tumor suppressor | 8.30E-02 | 6.80E-01
Cell motility | 9.00E-02 | 6.80E-01
Apoptosis | 9.40E-02 | 6.70E-01

**b**

```
dif. expressed (9/69)
random 1 (20/69)
random 2 (20/69)
```

**Predictor Model: KNN**

```
TRUE
```

---Training dataset:
Set A1 - KIRC (n=69, small RNAs)

---Test dataset:
Set A2 - KIRC (n=69, small RNAs)

**Figure S9.** Predictive power of differentially expressed exon 1 smRNAs in KIRC. a) Upper panel: analysis design. A similar approach was followed as in case of BRCA with the following modification. The set of Normal and Tumor KIRC samples were randomly divided into two new datasets (with an equal number of Normal/Tumor samples per dataset) (see Suppl. Table 4) and subsequently one of the two sets (setA1) was used to identify the differentially expressed smRNAs (see Suppl. Table 1). The respective GO term enrichment for genes at the identified smRNA features are presented at the lower panel, with terms marked in red being directly associated with cancer. b) Prediction model of the list of features in (a) compared to two randomly selected smRNA lists of the same length. The set that was used here as the test dataset was not the one used in (a) in order to prevent any positive bias when applying KNN. As in BRCA prediction potential of differentially expressed smRNAs surpasses that of random smRNA signatures. Numbers in parentheses correspond to misclassification rates. (for example 10/114 represents 10 misclassified samples out of 114 tested)