DNA methylation biomarkers discovered *in silico* detect cancer in liquid biopsies from non-small cell lung cancer patients

Vrba et al.

Supplemental materials:

Figure S1 …… 2

Figure S2 …… 3

Figure S3 …… 4

Figure S4 …… 5

Figure S5 …… 6

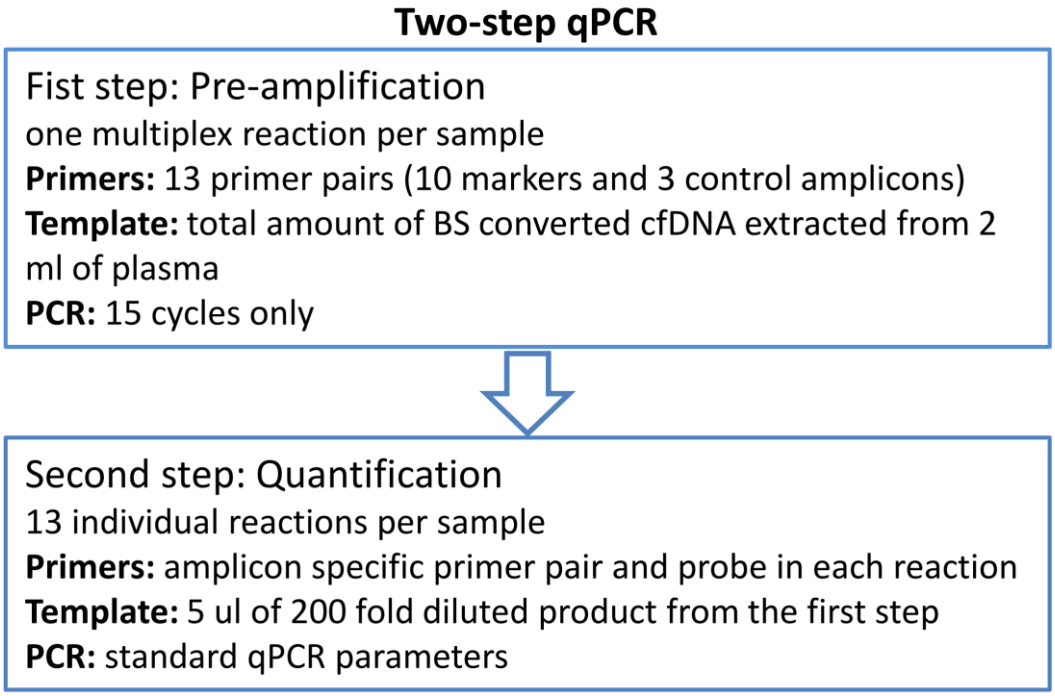
Figure S6 …… 7

Figure S7 …… 8

Table S1 …… 9

Table S2 …… 10

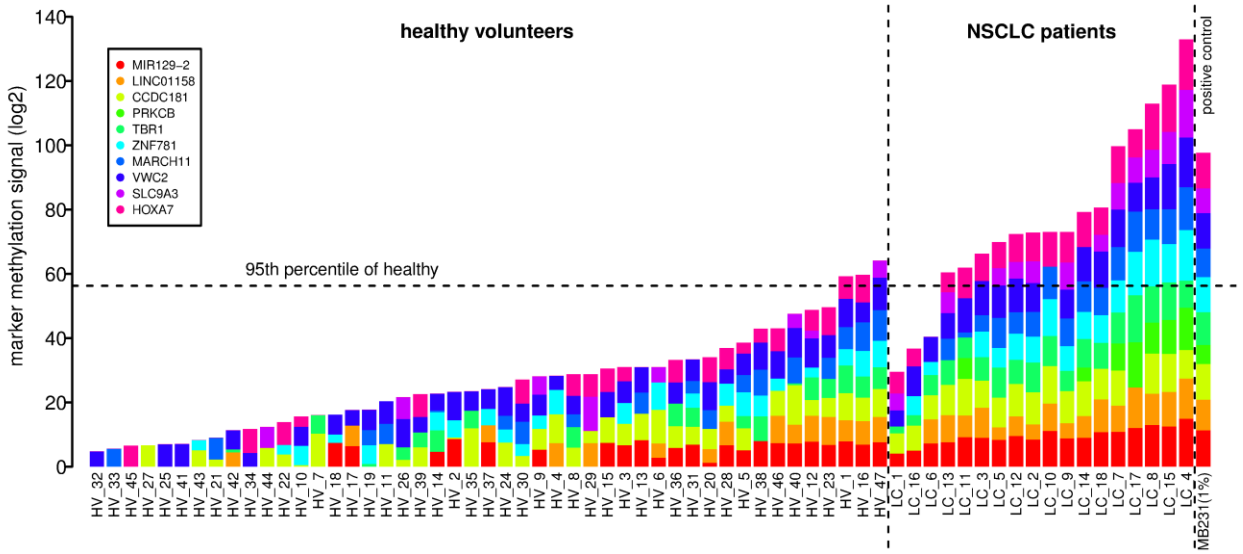
Figure S1



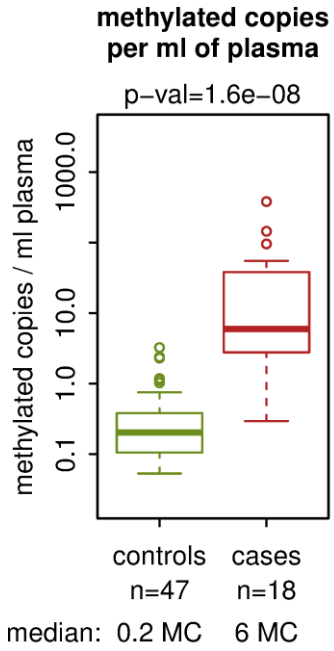
A schema of the two-step qPCR. First step: all methylated template molecules extracted from 2 ml of plasma are in contact with all primer pairs and therefore amplified. Second step: since all the available template was pre-amplified in the first step there is enough copies of each methylated marker to be representatively divided into individual marker specific reactions for quantification and therefore could be successfully detected even if the original amount was only several molecules.

Figure S2

A

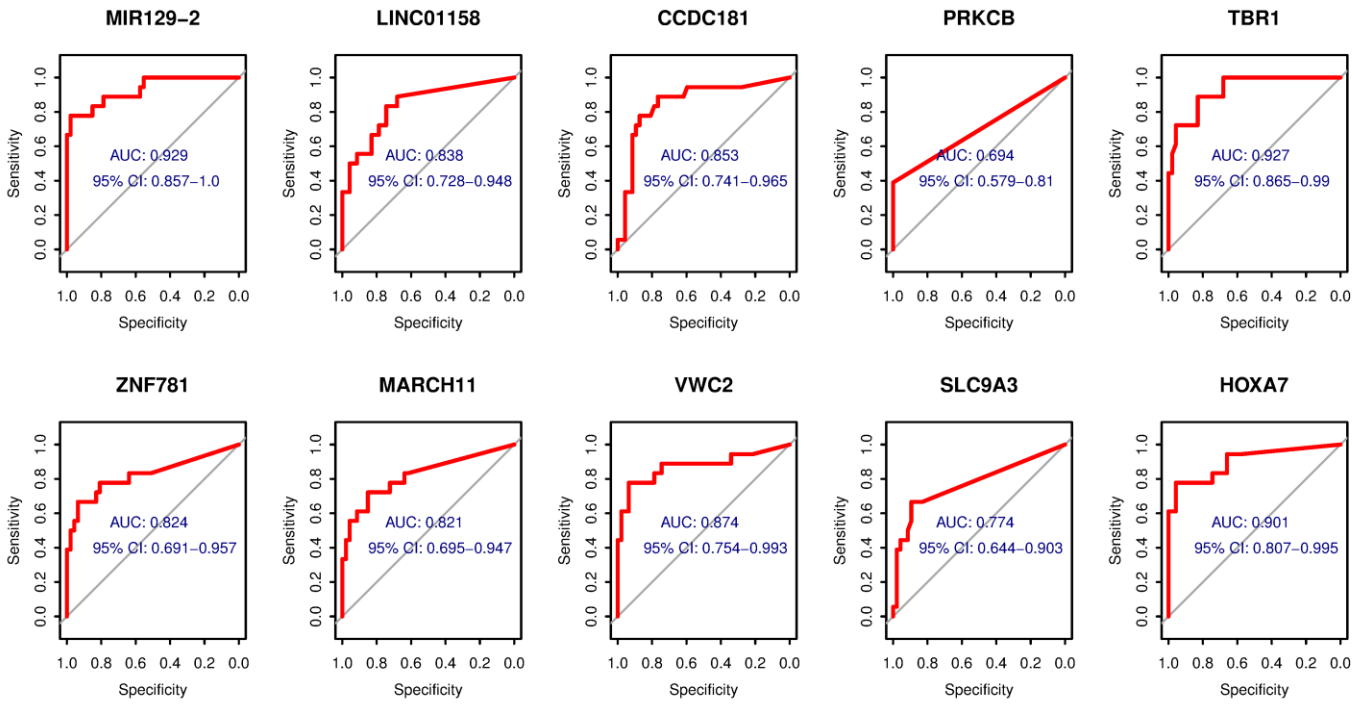


B



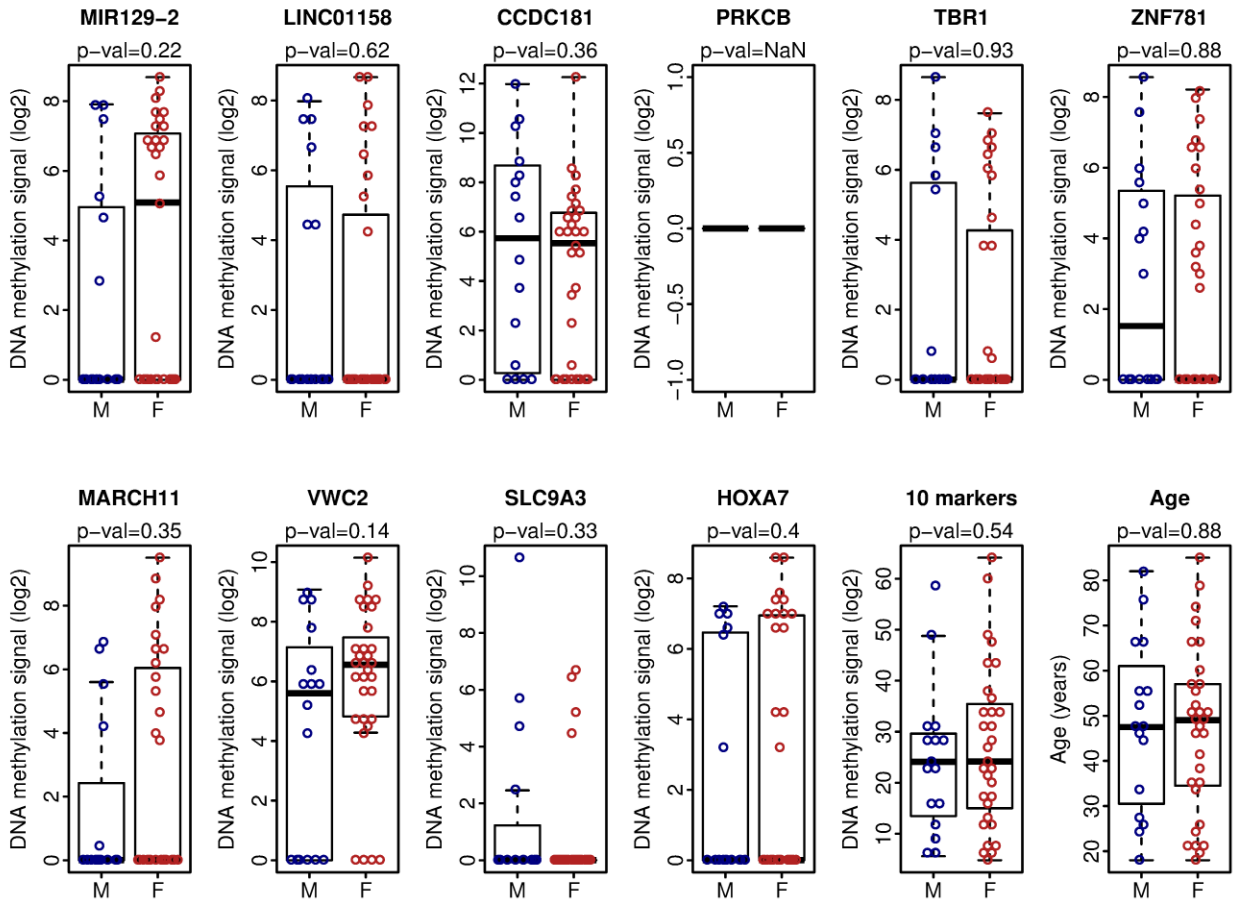
The panel A shows DNA methylation signal from the full 10 marker set on a cohort of 47 healthy subjects (left part) and 18 NSCLC patients (right part). The 95th percentile of the cumulative DNA methylation of the control cohort is represented by the horizontal dashed line. The panel B presents the estimated number of methylated template copies per marker per ml of plasma in both cohorts. The y axis is in a log scale. The median number of methylated template copies for each cohort is printed at the bottom. The amount of methylated copies per ml of plasma was estimated using the marker DNA methylation signal for each sample and the signal from control samples (normal human blood DNA spiked with DNA from MDA-MB231 cancer cells).

Figure S3



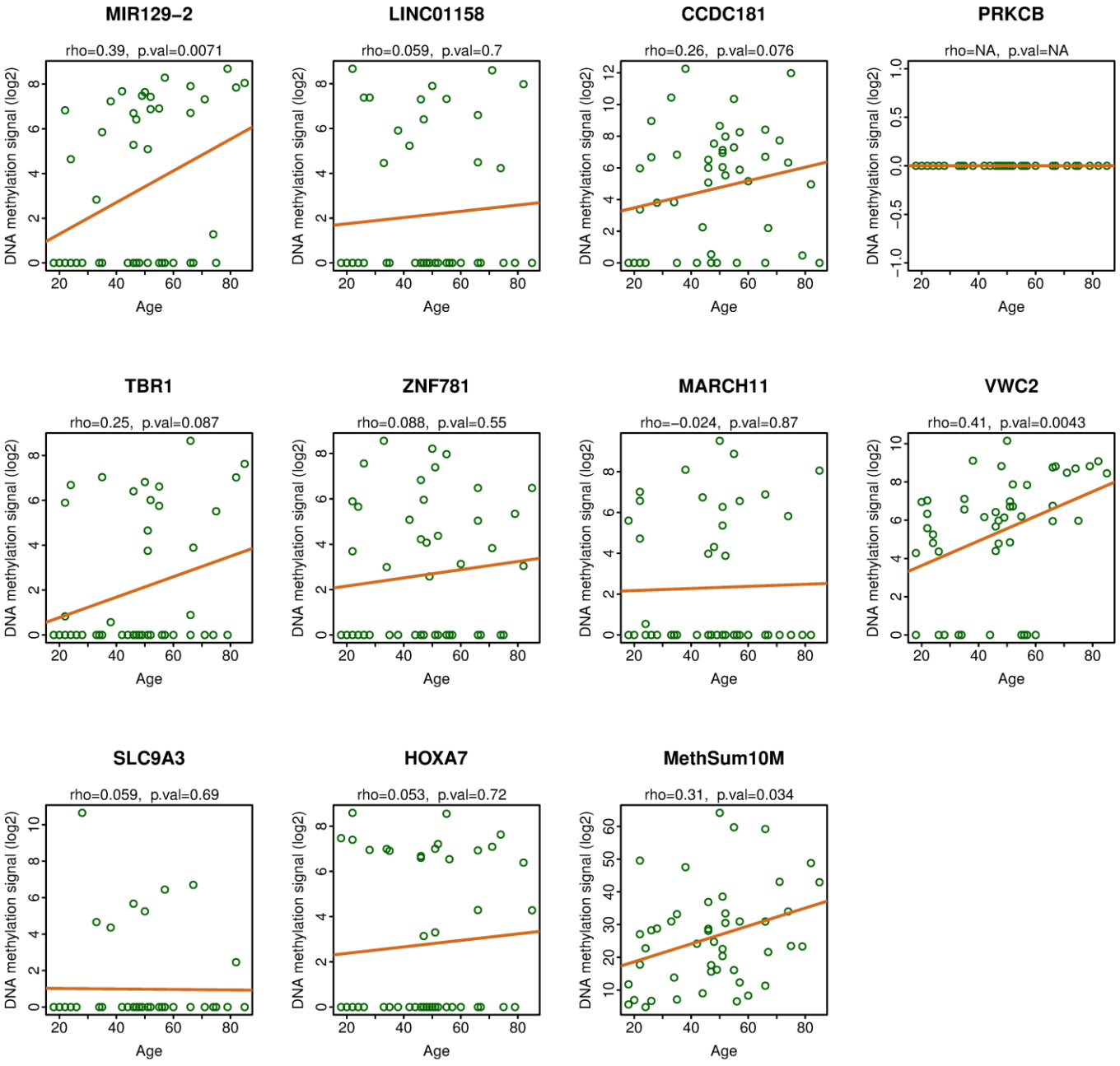
The performance of individual markers. ROC analysis of signal from individual markers using 18 NSCLC patients and 47 healthy subjects as control.

Figure S4



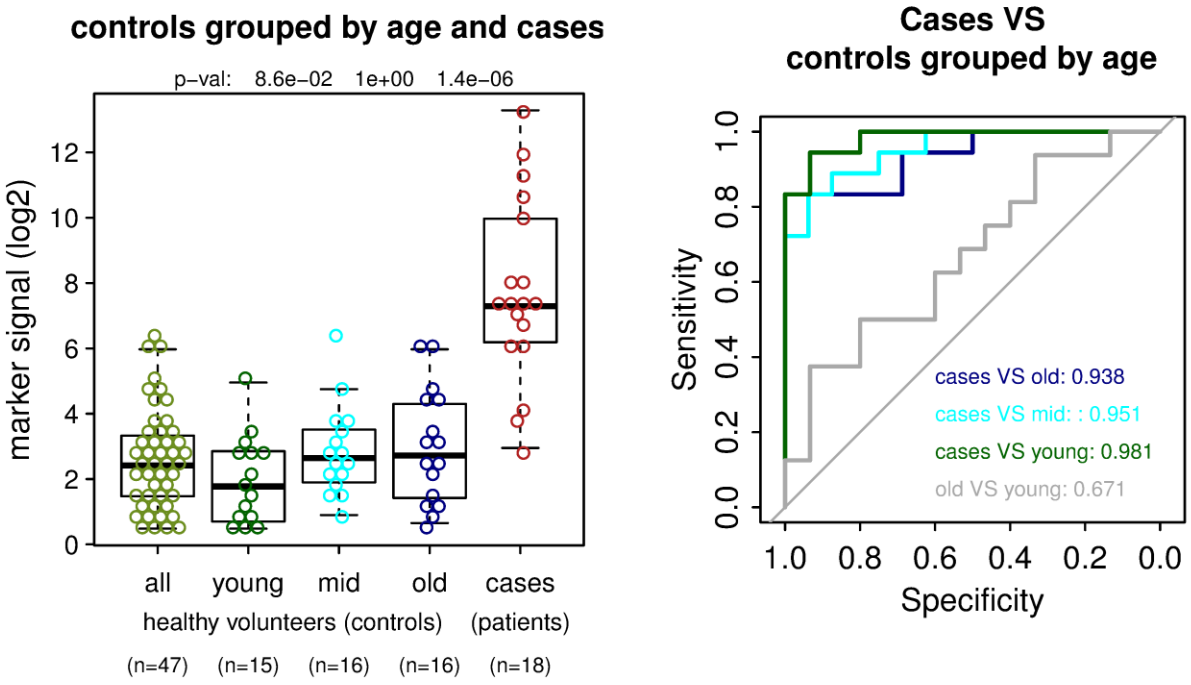
The analysis of DNA methylation signal of individual markers between sexes of healthy subjects. The first ten panels show data for individual markers. The last two panels show combined signal from all 10 markers and age, respectively. M = males, F = females.

Figure S5



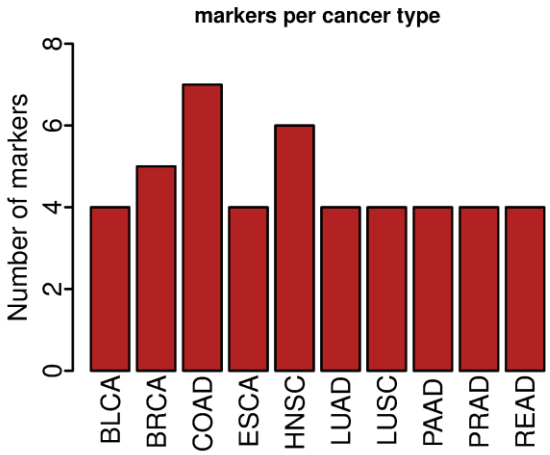
The relation between the DNA methylation signal from individual markers and the age of healthy subjects. The last panel shows the relation between the signal from the full marker set and age. The brown lines indicate the linear model fit. The Spearman correlation coefficients rho and the corresponding p-values are listed above each plot.

Figure S6



The DNA methylation signal of the 10 marker set in control samples grouped by age. The left panel shows DNA methylation signal in the entire control cohort, control cohort split into sub cohorts by age as describe in Fig 5A and in NSCLC cases cohort. The p-values for comparisons in marker signal between young and mid age controls, between mid age and old controls and between old controls and cases are listed at the top. The right panel shows that the marker set is able to distinguish between cases and individual age groups of controls with high sensitivity and specificity (blue, cyan and green curves) while the markers would be poor predictors to classify control samples into age groups (grey curve).

Figure S7



The numbers of markers out of the 10 marker set that passed all *in silico* filtering criteria[42] for individual TCGA cancer types.

Table S1

The description of the independent Illumina HumanMethylation450 DNA methylation data cohorts of cancer and normal tissue samples obtained from the GEO that were used to validate the marker set.

Tissue Name	TCGA cancer types represented	GEO accession	Place Of Data Origin	Count	Female	Male	Age Min	Age Max	Age Median	Data Level
bladder cancer	BLCA	GSE52955	PEBC, IDIBELL, Barcelona, Spain	25	NA	NA	NA	NA	NA	beta values
breast cancer	BRCA	GSE84207	Dept of Genetics, OUS Radiumhospitalet, Oslo, Norway	330	330	0	NA	NA	NA	IDAT
colorectal cancer	COAD, READ	GSE77718	Dept Biochemistry, University of Otago, Dunedin, New Zealand	96	44	49	NA	NA	NA	M and U signals
esophageal adenocarcinoma	ESCA	GSE72872	QIMR Berghofer MEdical Research Institute, Brisbane, QLD, Australia	125	NA	NA	NA	NA	NA	IDAT
tongue cancer	HNSC	GSE79556	Monash University, Clayton, Victoria, Australia	83	26	57	NA	NA	NA	beta values
NSCLC	LUAD, LUSC	GSE39279	PEBC, IDIBELL, Barcelona, Spain	444	190	254	35	90	66	beta values
pancreatic adenocarcinoma	PAAD	GSE49149	QIMR Berghofer MEdical Research Institute, Brisbane, QLD, Australia	167	73	94	34	90	69	IDAT
prostate cancer	PRAD	GSE84493	Ontario Institute for Cancer Research, Toronto, Ontario, Canada	202	0	202	NA	NA	NA	M and U signals
normal whole blood	-	GSE72773	University of California, Los Angeles, Los Angeles, CA, USA	310	150	160	35	91	69	M and U signals
normal bladder	-	GSE52955	PEBC, IDIBELL, Barcelona, Spain	5	NA	NA	NA	NA	NA	beta values
normal breast	-	GSE88883	Geisel School of Medicine at Dartmouth College, Hanover, NH, USA	100	100	0	18	82	37	IDAT
normal colorectal	-	GSE77718	University of Otago, Dunedin, New Zealand	96	45	49	NA	NA	NA	M and U signals
normal esophagus	-	GSE72872	QIMR Berghofer MEdical Research Institute, Brisbane, QLD, Australia	11	NA	NA	NA	NA	NA	IDAT
normal gingiva	-	GSE59962	Princess Margaret Cancer Centre, Toronto, Ontario, Canada	23	NA	NA	NA	NA	NA	beta values
normal lung	-	GSE52401	NCI/NIH, Rockville, MD, USA	244	36	208	NA	NA	NA	beta values
normal pancreas	-	GSE49149	QIMR Berghofer MEdical Research Institute, Brisbane, QLD, Australia	29	11	18	33	87	69	IDAT
normal prostate	-	GSE76938	HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA	63	0	63	44	73	61	M and U signals

Table S2

The description of the analytical amplicons including the amplicon sizes and primer and probe sequences. The bottom part shows genomic coordinates of the individual qPCR amplicons (hg19).

Name	Amplicon length	Chromosome	Forward primer	Reverse primer	Probe sequence
Markers:					
MIR129-2	70	11	GTTCGGTTTTAGGGTTCGGAGAT	CAAAATATACCGACTTCTTCGATTCG	Roche UPL70
LINC01158	86	2	TTTTATAGGGGTAGCGATTAGCGTTG	CTCTAAAACGCGCTCACCGAAA	TTTGGGTCGGGTTGGGTCGTTT
CCDC181	87	1	GGATATTGTATGCGTTTGCGTAGATT	CATAACAACAACGTACCTCTACGTCCTC	TCGTTTTCGTAGTTAGAGAGGTTCGGATG
PRKCB	71	16	CGGGCGAAGCGTACGGTGT	CGCAAAATAACTAACCCGACTACGA	Roche UPL70
TBR1	73	2	TGCGTTTTATCGATCGTACGTGTT	CCCGACTACGCTCCTCCGAC	Roche UPL70
ZNF781	78	19	GATTTAGTAGTCGTTGGTATAAGTTGCGT	CGATAAATCCGCGCACTCGAA	CGGAGACGTGGGAGCGTTTTTTTG
MARCH11	89	5	CGTTTCGGAATCGACGTGAGC	AAATTCGACTCCGAACGAACGA	TCGGTTCGTGGAGGCGGTT
VWC2	70	7	AGTGATAGGTTGGTTCGGCGTAGT	CTCGCGCTACCCCCGAAA	AACCCTACCGCCGCACCCGCT
SLC9A3	79	5	CGGTCGGTTACGTCGTCGAAT	CAACGAAACGAAAACGATTACGAA	CGTTATGGGTTTTTTTTCGTATTCGTATGT
HOXA7	68	7	TTGAGATTGGCGGAGGCGGTT	CCATTTTCTTTTAAACGAAACTCGC	TGTGGGCGGTTACGTGTTGCG
Controls:					
LRRC8A	81	9	TTGTATTTGACGGGTAATTTGAGCG	CTTAAAACGTTTAAACTCCCGCAAC	GGAGAATAATCGTTATATCGTTATCGACGG
NCOR2	74	12	GGGTTTTAGTTCGGAGCGGGT	GACCAAAACGACCCCGAACAA	TTTGGCGAGGAAGGTATGGTCGGT
TRAP1	68	16	GGTGACGGTTGGGGGCGTAT	AAAATACGCCAACCGCATACGA	GGTAGTAGATGTTGCGGGTGTCGGT

Name	Chromosome	Start	End
Markers:			
MIR129-2	chr11	43,602,876	43,602,945
LINC01158	chr2	105,459,225	105,459,310
CCDC181	chr1	169,396,658	169,396,744
PRKCB	chr16	23,847,491	23,847,561
TBR1	chr2	162,283,602	162,283,674
ZNF781	chr19	38,183,080	38,183,157
MARCH11	chr5	16,180,057	16,180,145
VWC2	chr7	49,813,047	49,813,116
SLC9A3	chr5	528,576	528,654
HOXA7	chr7	27,196,264	27,196,331
Controls:			
LRRC8A	chr9	131,671,004	131,671,084
NCOR2	chr12	124,821,510	124,821,583
TRAP1	chr16	3,714,325	3,714,392