

Table S1 Characteristics of the datasets used in this study

Datasets	Category	Sample size	Platform	Country
TCGA	Train set	50N + 491T	Illumina HumanMethylation450 BeadChip, Illumina HiSeq RNA-Seq	USA
GSE26126	Validation set	86N + 95T	Illumina HumanMethylation450 BeadChip	USA
GSE76938	Validation set	63N + 73T	Illumina HumanMethylation450 BeadChip	USA
GSE112047	Validation set	16N + 31T	Illumina HumanMethylation450 BeadChip	Canada
GSE84749	Validation set	4N + 20T	Illumina HumanMethylation450 BeadChip, Illumina HumanHT-12 V4.0 expression BeadChip	USA
Chaoyang cohort	Test set	20N + 20T	Illumina Hiseq	China

T = prostate cancer tissues, N = paracancerous tissues. TCGA, The Cancer Genome Atlas.