

Supplemental Table S1

Details of DNA methylation, gene expression and microRNA expression data from patient biopsies

Samples of the Illumina HumanMethylation450 arrays from 25 cancer datasets (23 distinct cancer types) used to identify differentially methylated regions (DMRs); gene expression data from RNA-seq, and microRNA data from smRNA-seq (except the GBM1 microRNA dataset derived from microarray). All data originated from patient biopsies.

#	Tumor	Abbrev.	Type of cancer	Methylation			Gene Expression			MiRNA Expression			Data Source	GEO accession #	
				T	N	Total	T	N	Total	T	N	Total		Methylation	Gene expression
1	B-Cell lymphoma	BCELL	P	18	56	74	4	5	9	0	0	0	GEO	GSE37362 GSE59250	GSE37363
2	bladder urothelial carcinoma	BLCA	P	191	20	211	182	18	200	273	19	292	TCGA		
3	breast invasive carcinoma	BRCA	P	684	98	782	998	108	1106	1066	103	1169	TCGA		
4	cholangiocarcinoma	CHOL	P	36	9	45	0	0	0	0	0	0	TCGA		
5	colon adenocarcinoma	COAD1	P	292	38	330	233	21	254	429	8	437	TCGA		
6	colon adenocarcinoma	COAD2	P	23	40	63	0	0	0	0	0	0			
7	glioblastoma multiforme	GBM1	P	125	44	169	161	0	161	571	10	581	TCGA & GEO	GSE36278 GSE41826 GSE50798	
8	glioblastoma multiforme	GBM2	P	198	6	204	0	0	0	0	0	0			
9	head & neck squamous cell carcinoma	HNSC	P	373	50	423	303	37	340	513	44	557	TCGA		
10	kidney clear cell carcinoma	KIRC	P	309	160	469	0	0	0	535	71	606	TCGA		
11	kidney papillary cell carcinoma	KIRP	P	142	45	187	141	30	171	242	34	276	TCGA		
12	brain lower grade glioma	LGG	P	266	2	268	271	0	271	518	0	518	TCGA & GEO	GSE50798	
13	hepatocellular adenomas	LIVER	P	50	4	54	0	0	0	0	0	0	GEO	GSE43091	
14	lung adenocarcinoma	LUAD	P	452	32	484	468	58	526	499	46	545	TCGA		
15	lung squamous cell carcinoma	LUSC	P	361	43	404	483	50	533	467	45	512	TCGA		
16	myelodysplastic syndrome	MDS	BP	4	4	8	8	8	16	0	0	0	GEO	GSE51758	GSE51759
17	oral squamous cell carcinoma	OSCC	P	42	2	44	0	0	0	0	0	0	GEO	GSE41114	
18	pilocytic astrocytoma	PA	BP	61	6	67	0	0	0	0	0	0	GEO	GSE44684	
19	prostate adenocarcinoma	PRAD	P	217	49	266	195	45	240	425	52	477	TCGA		
20	rectum adenocarcinoma	READ	P	96	7	103	84	5	89	159	3	162	TCGA		
21	skin cutaneous melanoma	SKCM1	P	71	4	75	68	3	71	81	2	83	TCGA		GSE38495
22	stomach adenocarcinoma	STAD	P	261	2	263	249	33	282	382	43	425	TCGA		
23	thymoma	THYM	P	124	2	126	0	0	0	0	0	0	TCGA		
24	thyroid carcinoma	THCA	P	508	56	564	494	58	552	507	59	566	TCGA		
25	uterine corpus endometrioid carcinoma	UCEC	P	415	46	461	118	17	135	542	33	575	TCGA		
26	metastatic melanoma to adrenal gland	Adrenal	MM	2	0	2	0	0	0	0	0	0	TCGA		
27	metastatic melanoma to bone	Bone	MM	1	0	1	0	0	0	0	0	0	TCGA		
28	metastatic melanoma to brain	Brain	MM	17	0	17	0	0	0	0	0	0	TCGA & GEO	GSE44661	
29	metastatic melanoma to gastrointestinal tract	Gastro	MM	6	0	6	0	0	0	0	0	0	TCGA		
30	metastatic melanoma to lymph node	Lymph	MM	8	0	8	0	0	0	0	0	0	TCGA		
31	metastatic melanoma to lung	Lung	MM	4	0	4	0	0	0	0	0	0	TCGA		
32	metastatic melanoma to other subcutaneous areas, fat and muscle	Subcut	MM	16	0	16	0	0	0	0	0	0	TCGA		
Totals				5373	825	6198	4460	496	4956	7209	572	7781			

Notes:

All data was derived from patients samples

MiRNA expression data was derived from small RNA-Seq experiments (except GBM which is microRNA microarray)

Data source: All data was derived from TCGA and GEO databases, if GEO is states in the 'Src' column, the GEO accession code is provided in the specified column

Abbreviations (Abbrev.):

N	Normal samples
T	Tumor samples
miRNA	gene expression data
miRNA	microRNA
TCGA	The Cancer Genome Atlas, https://tcga-data.nci.nih.gov/tcga/
GEO	Gene Expression Omnibus (GEO) repository, http://www.ncbi.nlm.nih.gov/gds/

The 'Type of cancer column' refers only to the cancer samples, not to the normal samples, abbreviations are as follows:

BP	Benign primary tumor
P	Primary tumor samples
MM	Metastatic melanoma