

RRBS Dataset From Richard Myers, HAIB (Encode Project)

39 data from 16 tissues in ENCODE

(samples for each tissue \geq 3)



14 data from 7 tissues in ENCODE

(samples for each tissue \geq 3)

ENCFF000LVO	RRBS	lung	adult	female	83 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVR	RRBS	lung	adult	female	83 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LUT	RRBS	breast	adult	female	21 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LUV	RRBS	breast	adult	female	21 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LUQ	RRBS	brain	adult	male	66 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LUU	RRBS	brain	adult	male	66 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LWS	RRBS	stomach	adult	female	83 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LWW	RRBS	stomach	adult	female	83 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVU	RRBS	pancreas	adult	male	71 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVW	RRBS	pancreas	adult	male	71 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVI	RRBS	mononuclear cell	adult	unknown	54 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVK	RRBS	mononuclear cell	adult	unknown	54 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVJ	RRBS	liver	adult	female	83 year	36bp	Illumina Genome Analyzer IIX
ENCFF000LVN	RRBS	liver	adult	female	83 year	36bp	Illumina Genome Analyzer IIX

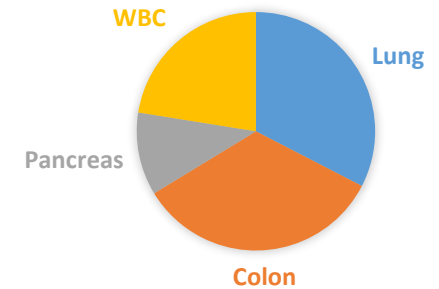
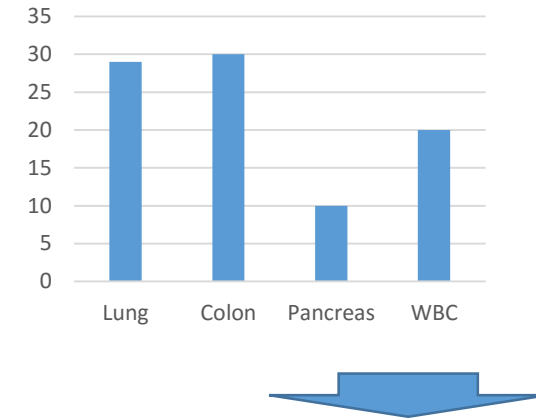
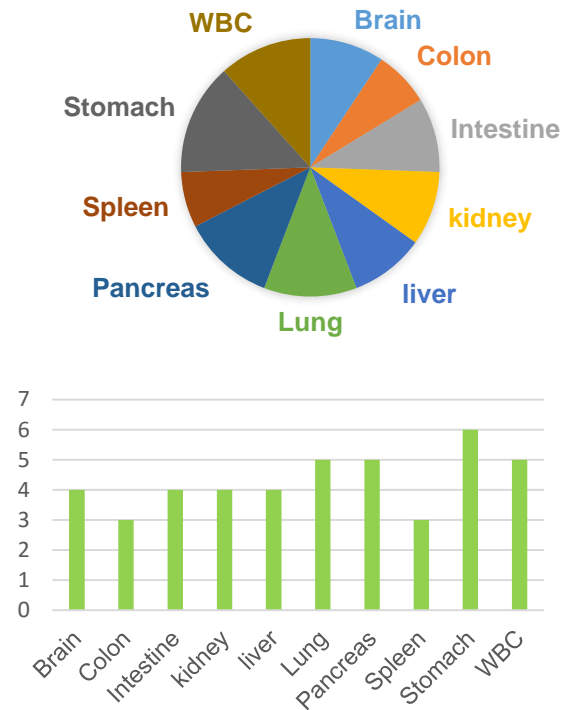
Prediction for Full tissues based on Random Forest Model

43 samples from 10 tissues

(samples for each tissue>=3)

Salk: 18 N37: 7 WBC: 3 ENCODE: 14

Sample ID	Format	Tissue	Lab
ENCFF000LUQ	RRBS	Brain	Haib
ENCFF000LUU	RRBS	Brain	Haib
N37-Cerebellum (CRBL)	WGBS	Brain	N37
N37-Frontal lobe (FL)	WGBS	Brain	N39
STL001SG-01	WGBS	Colon	Salk
STL003SG-01	WGBS	Colon	Salk
N37-Colon	WGBS	Colon	N38
STL001SB-01	WGBS	Intestine	Salk
STL002SB-01	WGBS	Intestine	Salk
STL003SB-01	WGBS	Intestine	Salk
N37-Small intestine (SI)	WGBS	Intestine	N41
ENCFF000LVA	RRBS	kidney	Haib
ENCFF000LVB	RRBS	kidney	Haib
STL002AD-01	WGBS	Kidney	Salk
STL003AD-01	WGBS	Kidney	Salk
ENCFF000LVJ	RRBS	liver	Haib
ENCFF000LVN	RRBS	liver	Haib
STL011LI-01	WGBS	Liver	Salk
N37-Liver	WGBS	Liver	N42
ENCFF000LVO	RRBS	Lung	Haib
ENCFF000LVR	RRBS	Lung	Haib
STL001LG-01	WGBS	Lung	Salk
STL002LG-01	WGBS	Lung	Salk
N37-Lung	WGBS	Lung	N43
ENCFF000LVU	RRBS	Pancreas	Haib
ENCFF000LVW	RRBS	Pancreas	Haib
STL002PA-01	WGBS	Pancreas	Salk
STL003PA-01	WGBS	Pancreas	Salk
N37-Pancreas	WGBS	Pancreas	N45
STL001SX-01	WGBS	Spleen	Salk
STL002SX-01	WGBS	Spleen	Salk
STL003SX-01	WGBS	Spleen	Salk
STL001GA-01	WGBS	Stomach	Salk
STL003GA-01	WGBS	Stomach	Salk
ENCFF000LWS	RRBS	Stomach	Haib
ENCFF000LWW	RRBS	Stomach	Haib
STL002GA-01	WGBS	Stomach	Salk
N37-Stomach	WGBS	Stomach	N46
ENCFF000LVI	RRBS	WBC	Haib
ENCFF000LVK	RRBS	WBC	Haib
WB_centenarian.mld_blocks_r2-0.5	WBC	WBC	Age
WB_middle-age.mld_blocks_r2-0.5	WBC	WBC	Age
WB_new-born.mld_blocks_r2-0.5	WBC	WBC	Age



	Colon	Lung	Pancreas	WBS
Brain	0.243	0.159	0.3	0.149
Colon	21.06	0.188	0.17	0.213
Intestine	2.398	0.126	0.234	0.136
kidney	0.811	0.946	0.632	0.079
liver	0.373	0.263	0.997	0.31
Lung	0.173	25.577	0.875	0.576
Pancreas	0.643	0.232	6.324	0.756
Spleen	0.935	1.067	0.231	0.245
Stomach	1.222	0.542	0.209	0.46
WBC	0.742	0.6	0.18	17.075
Sensitivity	0.726	0.852	0.632	0.853

10000 iteration, 49 predictors

95% confidence interval will be calculated after all the iteration finished.

Next step: check these 49 predictors, differential methylated?

Gene summaries powerful in the mapping prediction

1	chr2	10281508	10351856	10282009	C2orf48
2	chr2	10443029	10567743	10559883	HPCAL1
3	chr1	10696665	10856733	10698998	CASZ1
4	chr19	10828728	10942586	10828918	DNM2
5	chr19	10982252	11033448	10982378	CARM1
6	chr16	11038344	11276046	11038574	CLEC16A
7	chr6	11183530	11382581	11382581	NEDD9
8	chr16	11762288	11773015	11769915	SNN
9	chr19	13106583	13209610	13106651	NFIX
10	chr19	13317255	13617274	13318859	CACNA1A
11	chr10	13319795	13341746	13320300	PHYH
12	chr10	13685705	14372866	13693904	FRMD4A
13	chr19	13906273	13943044	13906462	ZSWIM4
14	chr19	15225784	15236610	15226062	ILVBL
15	chr19	15562437	15575382	15562605	RASAL3
16	chr1	16450831	16482582	16451709	EPHA2
17	chr20	17594322	17662928	17594826	RRBP1
18	chr17	17584786	17714765	17696262	RAI1
19	chr19	17935592	17958841	17937551	JAK3
20	chr19	18263987	18281343	18266689	PIK3R2
21	chr22	18270415	18507325	18273497	MICAL3
22	chr19	18979354	19006953	18979405	GDF1
23	chr19	18979354	19006953	18981385	CERS1
24	chr22	19701986	19710845	19702111	SEPT5
25	chr22	19704742	19712297	19712297	SEPT5-GP1BB
26	chr18	19749397	19782491	19751105	GATA6
27	chr20	19867164	19983103	19867313	RIN2
28	chr22	19929262	19957498	19950049	COMT
29	chr16	22430866	22449036	22449036	RRN3P3

* Variable importance(done)

30	chr8	22570764	22785421	22570882	PEBP4
31	chr22	23401592	23484241	23401639	RSPH14
32	chr22	23412668	23467221	23437882	GNAZ
33	chr2	23608297	23931483	23785066	KLHL29
34	chr22	24407764	24574596	24431965	CABIN1
35	chr1	25226001	25291501	25228612	RUNX3
36	chr10	26727265	26856732	26781255	APBB1IP
37	chr7	27145808	27166639	27147533	HOXA3
38	chr16	27214806	27233089	27221444	KDM8
39	chr1	27668482	27680423	27671825	SYTL1
40	chr22	28144264	28197486	28146902	MN1
41	chr16	28565248	28603111	28592390	SGF29
42	chr16	30125425	30134630	30127988	MAPK3
43	chr17	30819539	31204191	30821776	MYO1D
44	chr16	31483475	31489281	31483509	TGFB1I1
45	chr3	32859509	32933771	32859572	TRIM71
46	chr14	32798478	33302268	32902699	AKAP6
47	chr6	34857037	35059190	34857179	ANKS1A
48	chr6	35441373	35464861	35442820	TEAD3
49	chr17	38465422	38513895	38487470	RARA

* Differential analysis

*Several Figure to show the methylation status

*All the GSI>0.5

* Few of them were reported in previous study

* More Analysis to these genes need to be done