

WAS cell lineage study - supplemental 1

John K. Everett

Full list of oncogenes considered (2,027 genes)

When considering all subjects, a Fisher's exact test of unique intSites near oncogenes, comparing first time points (27.79% intSites near oncogenes) and second time points (27.01% intSites near oncogenes), showed a small yet significant depletion of intSites near oncogenes in the second time points (p-value: 0.0004).

	nSites > 50 KB oncogenes	nSites \leq 50 KB oncogenes
First time point	94,494	36,357
Second time point	43,807	16,211

For each of the six subjects, comparing the percentages of intSites near oncogenes from the first and second time points showed no significant difference (paired Wilcoxon test, p-value: 0.844).

Subject	TP1 percent sites near oncogenes	TP1 sites near / far from oncogenes	TP2 percent sites near oncogenes	TP2 sites near / far from oncogenes
b0/bE	28.7%	22,999 / 57,133	28.29%	7,265 / 18,412
bS/bS	29.36%	1,582 / 3,806	29.17%	1,362 / 3,307
WAS2	26.48%	3,081 / 8,554	26.95%	3,125 / 8,470
WAS4	25.89%	7,161 / 20,503	22.57%	2,017 / 6,918
WAS5	27.51%	1,006 / 2,651	25.04%	601 / 1,799
WAS7	22.23%	528 / 1,847	27.31%	1,841 / 4,901

Lymphoma associated oncogenes considered (38 genes)

When considering all subjects, a Fisher's exact test of unique intSites near oncogenes, comparing first time points (0.65% intSites near oncogenes) and second time points (0.65% intSites near oncogenes), no significant difference was found (p-value: 1.00).

	nSites > 50 KB oncogenes	nSites \leq 50 KB oncogenes
First time point	130,001	850
Second time point	59,628	390

For each of the six subjects, comparing the percentages of intSites near oncogenes from the first and second time points showed no significant difference (paired Wilcoxon test, p-value: 0.563).

Subject	TP1 percent sites near oncogenes	TP1 sites near / far from oncogenes	TP2 percent sites near oncogenes	TP2 sites near / far from oncogenes
b0/bE	0.69%	554 / 79,578	0.74%	190 / 25,487
bS/bS	0.65%	35 / 5,353	0.64%	30 / 4,639
WAS2	0.6%	70 / 11,565	0.58%	67 / 11,528
WAS4	0.59%	163 / 27,501	0.46%	41 / 8,894
WAS5	0.57%	21 / 3,636	0.75%	18 / 2,382
WAS7	0.29%	7 / 2,368	0.65%	44 / 6,698

The table below shows the controls and results of those controls for each sequencing run used in this study. Four technical replicates were employed for each control experiment. Two intSites were found in the uninfected control experiment for sequencing run 160426_M03249_0159_000000000-AM2PA which included a single study sample (WAS7, m12/GRANULOCYTES, 879 intSites).

Sequencing experiment controls

Run	Number of samples	Uninfected control	Uninfected control sites	No DNA control	No DNA control sites
150218_M03249_0040_000000000-ADTJB	4	Yes	0	No	NA
150505_M03249_0067_000000000-AFCKK	5	Yes	0	No	NA
150716_M03249_0081_000000000-AGRLF	4	Yes	0	No	NA
150903_M03249_0094_000000000-AJ0VV	4	Yes	0	No	NA
151116_M00142_0033_000000000-AGL5A	2	Yes	0	No	NA
151210_M03249_0115_000000000-AKW5U	5	Yes	0	No	NA
151211_M03249_0116_000000000-AL0NJ	5	No	NA	No	NA
151214_M03249_0117_000000000-AKW0A	5	Yes	0	No	NA
151215_M03249_0118_000000000-AK08F	6	No	NA	No	NA
160426_M03249_0159_000000000-AM2PA	1	Yes	2	No	NA
160511_M03249_0163_000000000-APCV7	5	Yes	0	Yes	0
160825_M03249_0186_000000000-AT6V3	4	Yes	0	Yes	0
161107_M03249_0205_000000000-AW45T	8	Yes	0	Yes	0
170216_M03249_0228_000000000-AVFW2	1	Yes	0	Yes	0
170421_M03249_0246_000000000-AYHME	10	Yes	0	Yes	0