BAP1 KO/KI Experiments

Joseph McElroy

12 May, 2020

Contents

Objective

Identify associations between human and mouse samples in various BAP1 statuses.

Request

Email

"Attached is the RNAseq results of the BAP1 knock out and knock in experiments for one mouse melanoma, one human liver and one human lung cell lines. These are what Maciej Pietrzak sent us

The three B16 are mouse in origin. The B16 is the original cell line B16 2E10 and B16 4F7 are two BAP1 KO clones (CRISPR) from the original

The SNU and NCI are human in origin

The SNU is the original cell line. The SNU 2D3 and SNU 1F5 are two different KO clones (CRISPR)

The NCI is a BAP1 null cell line with biallelic large deletions of BAP1 we transfected it with either empty vector (EV) or full length wild type BAP1. There is only one replica of this."

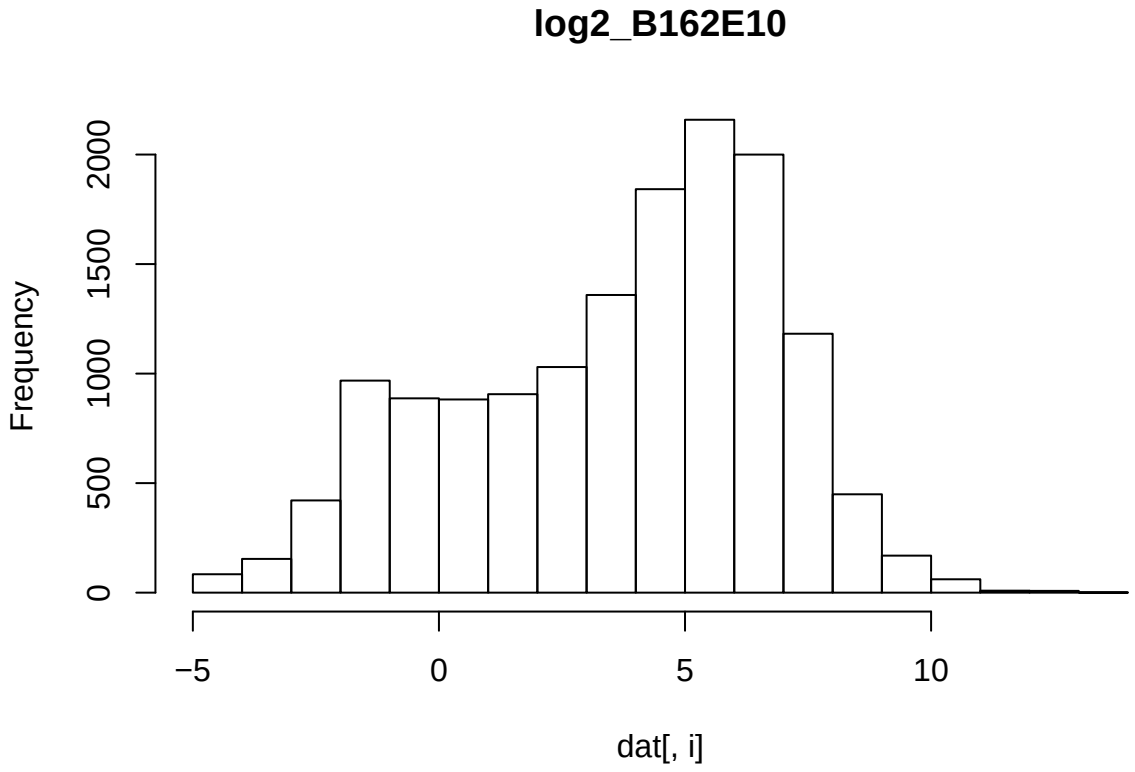
Data

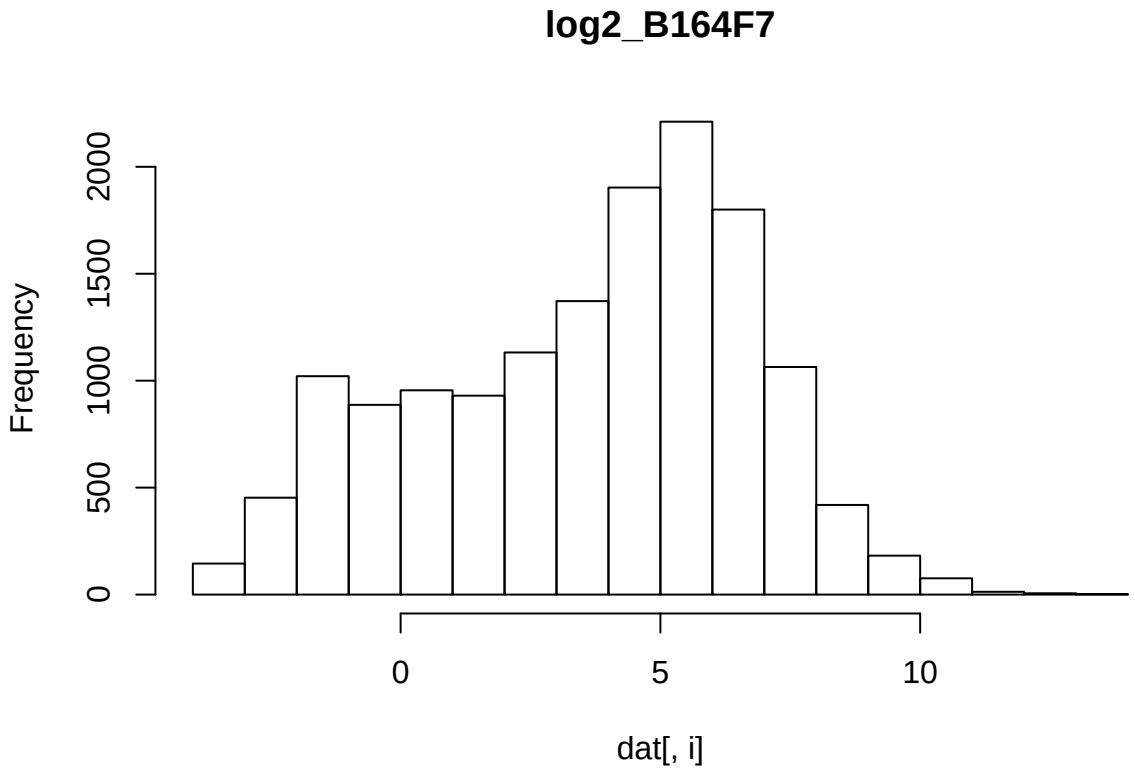
Data are in counts per million.

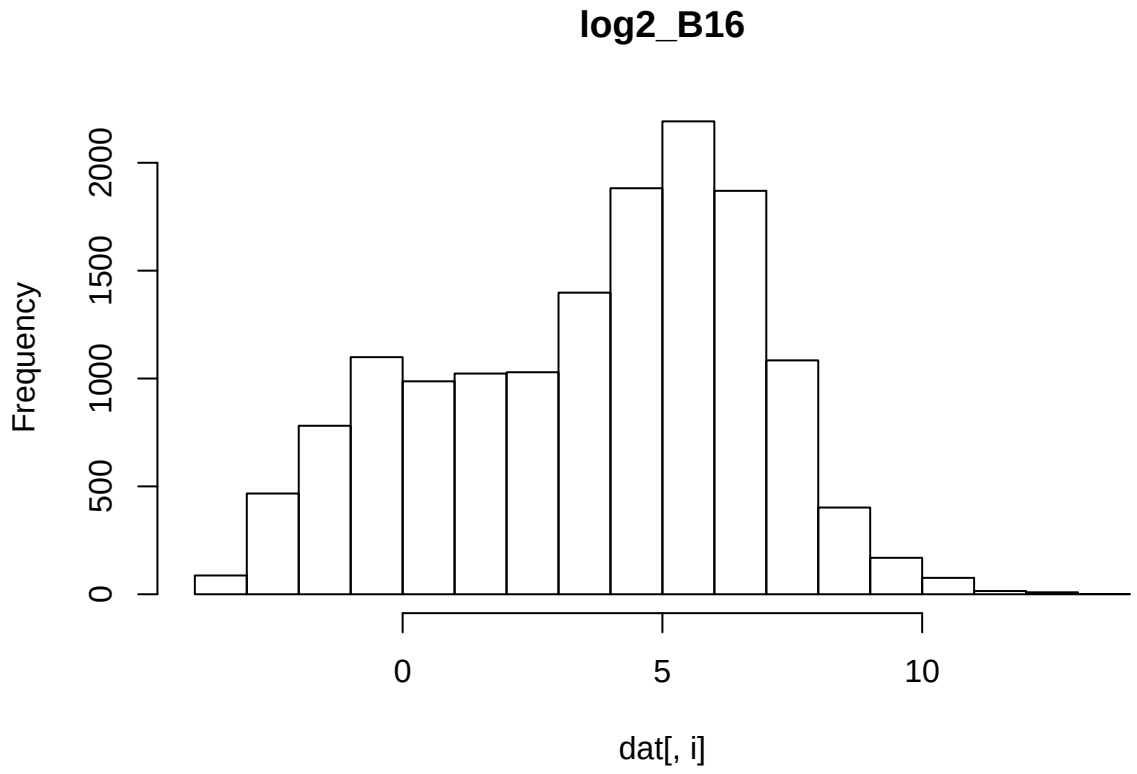
Distributions

Will compare distributions of log2 counts.

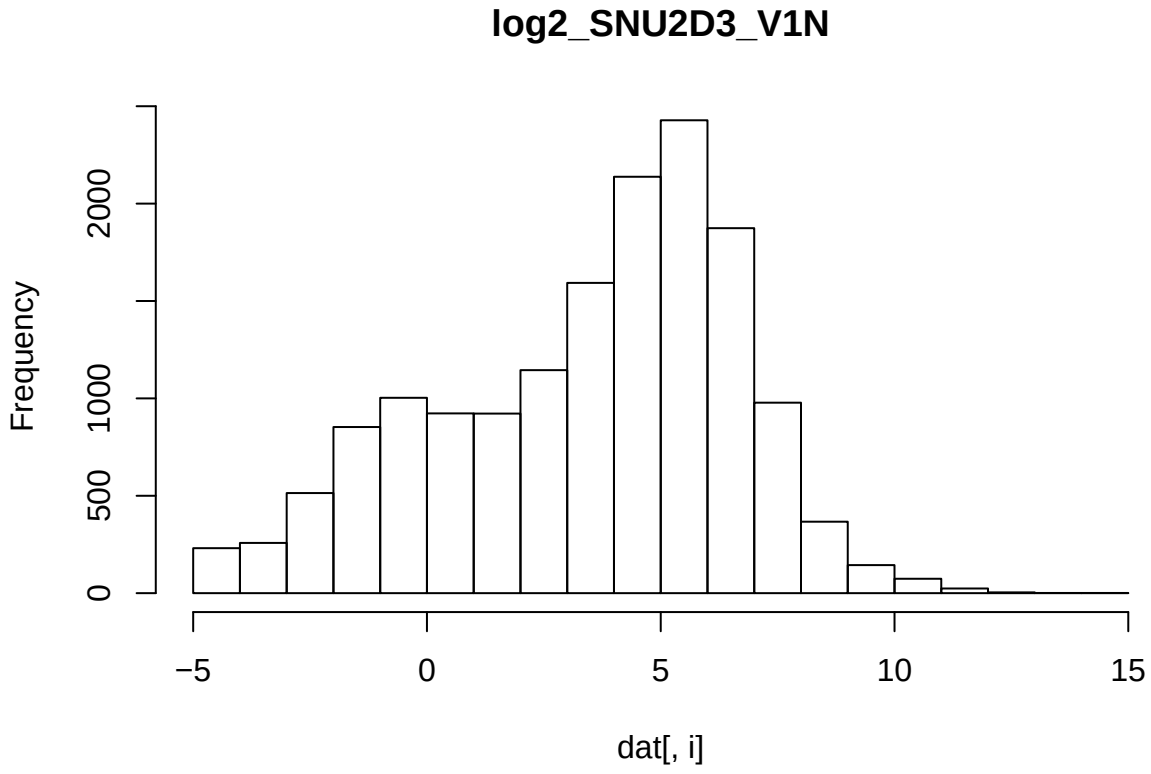
Mouse

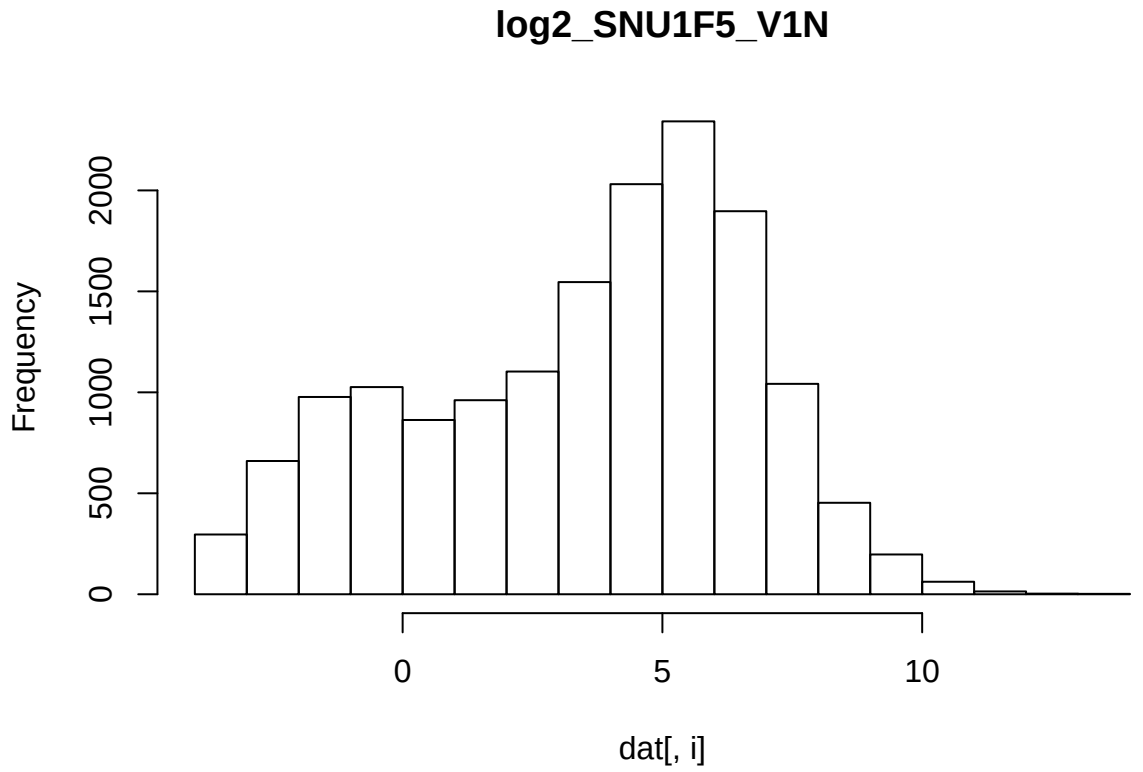

log2_B162E10
Frequency
0
500
1000
1500
2000
−5
0
5
10
dat[, i]

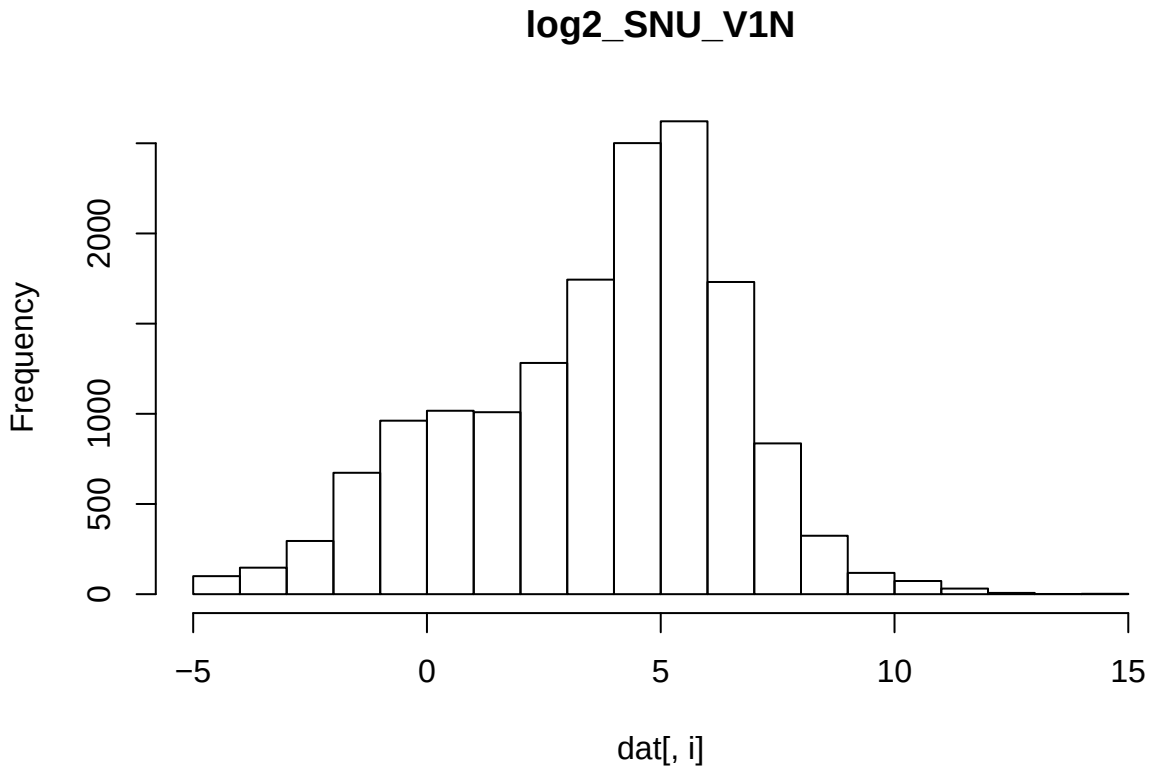

log2_B164F7
Frequency
0
500
1000
1500
2000
0
5
10
dat[, i]


log2_B16
Frequency
0
500
1000
1500
2000
0
5
10
dat[, i]

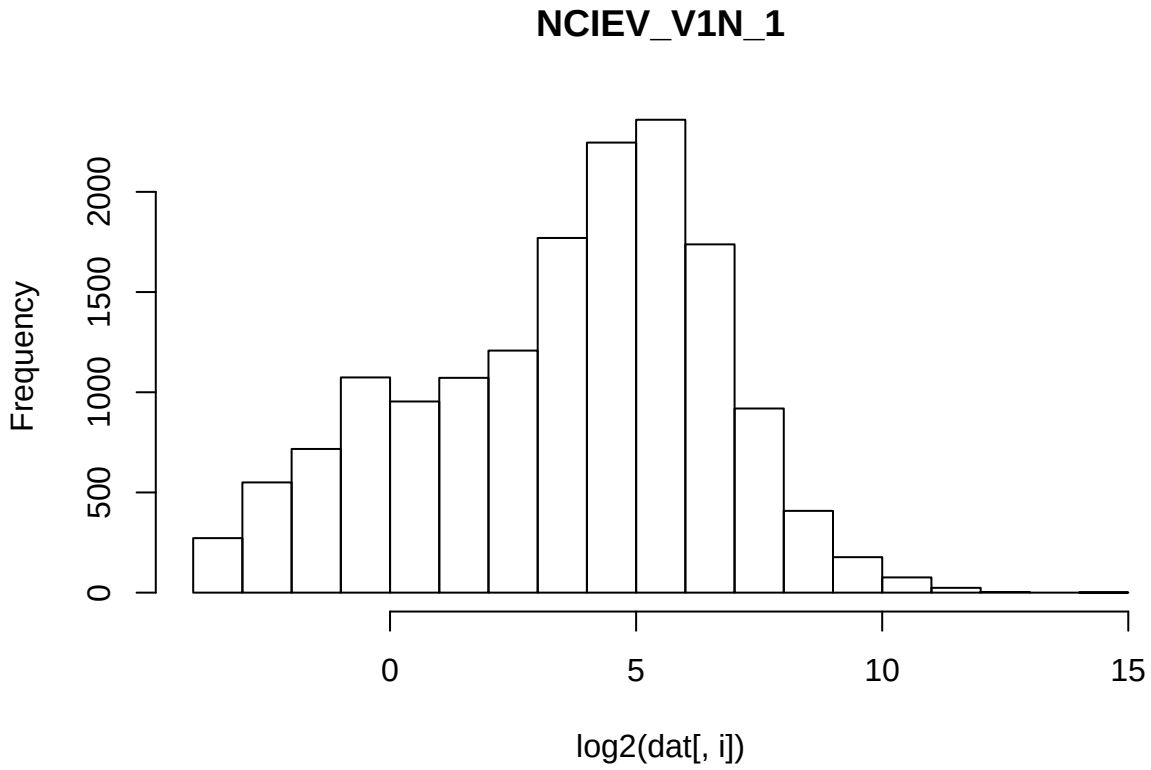
Human snu

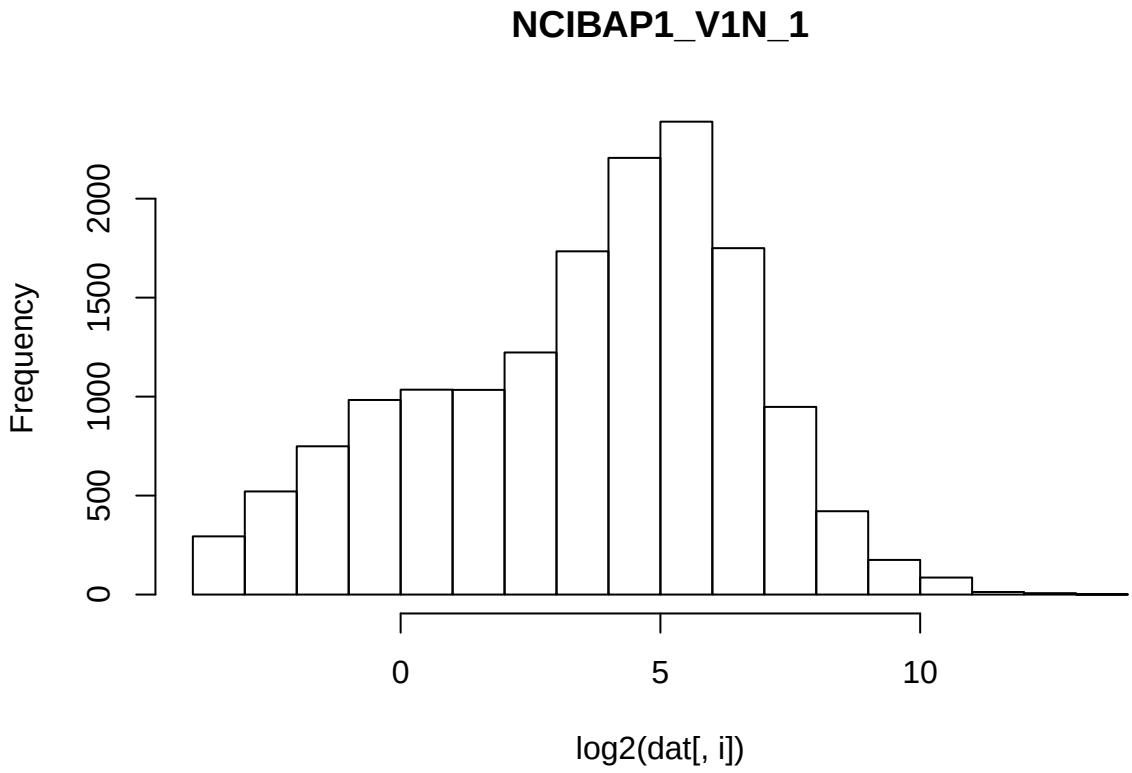

log2_SNU2D3_V1N
Frequency
0
500
1000
2000
−5
0
5
10
15
dat[, i]


log2_SNU1F5_V1N
Frequency
0
500
1000
1500
2000
0
5
10
dat[, i]


log2_SNU_V1N
Frequency
0
500
1000
2000
−5
0
5
10
15
dat[, i]

Human NCI





Distributions across all samples within a group look very similar, so will not perform additional normalization.

Methods

Data in Log2 space, except for the NCI, because a different type of analysis is done for it. For the comparisons where there are 2 vs. 1, a linear model (t-test assuming equal variance) is employed. For the comparison of 1 vs 1, a normal distribution will be fit to the log2 ratio of human mutant vs mutant to get a null distribution, and then the genes from the NCI $< 2.5\%$ lowest and $> 2.5\%$ highest are selected. Under this distribution, we expect 5% of comparisons to be significant by chance alone, so the overall FDR of the method will be calculated as (5%*total comparisons)/number of declared associations.

Results

Mouse Data

WT is reference.

Top "hits":

Ensembl.ID	Gene.Name	estimate	raw_pvalue	FDR_q
ENSMUSG00000106106	CT010467.1	-0.3075	7.259e-05	0.9701
ENSMUSG00000050891	Tatdn1	-0.5964	0.0001783	0.9701
ENSMUSG00000015839	Nfe2l2	0.3554	0.0002993	0.9701
ENSMUSG00000052384	Nrros	0.8872	0.0006282	0.9701
ENSMUSG00000030110	Ret	-1.283	0.0006738	0.9701
ENSMUSG00000031168	Ebp	-0.1108	0.0006906	0.9701
ENSMUSG00000025470	Zfp511	0.2193	0.0007718	0.9701
ENSMUSG00000074024	4632427E13Rik	-1.007	0.0008583	0.9701
ENSMUSG00000042750	Bex2	0.9191	0.0009403	0.9701
ENSMUSG00000030741	Spns1	0.6182	0.0009686	0.9701

Full results in:

"B16_res.xlsx"

Human Data

SNU

WT is reference.

Top "hits":

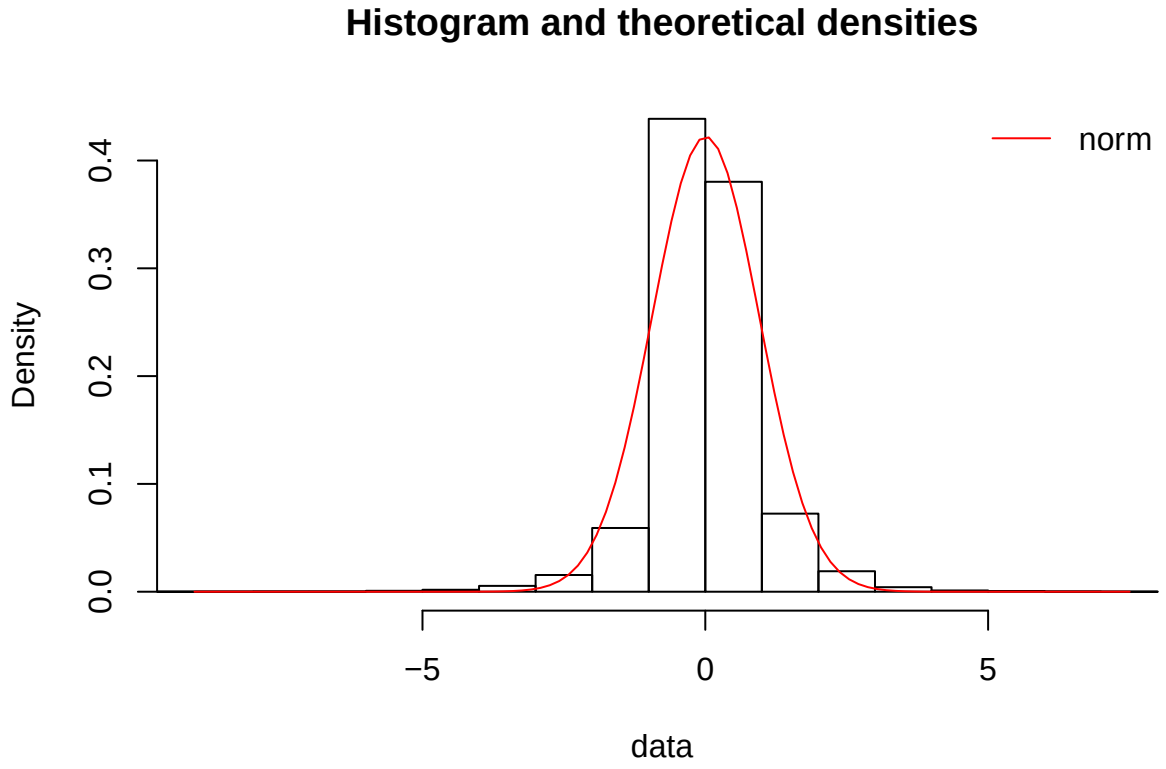
Ensembl.ID	Gene.Name	estimate	raw_pvalue	FDR_q
ENSG00000114859	CLCN2	0.5807	2.699e-05	0.2951
ENSG00000126001	CEP250	0.7821	4.281e-05	0.2951
ENSG00000096872	IFT74	-0.7466	5.72e-05	0.2951
ENSG00000205339	IPO7	0.4925	9.195e-05	0.3557
ENSG00000213339	QTRT1	0.4246	0.0002651	0.6418
ENSG00000214706	IFRD2	1.158	0.0003901	0.6418
ENSG00000133135	RNF128	-3.759	0.0003993	0.6418
ENSG00000185236	RAB11B	0.7009	0.0004437	0.6418
ENSG00000267278	MAP3K14-AS1	-0.5937	0.0005135	0.6418
ENSG00000184357	H1-5	0.5148	0.0005922	0.6418

Full results in:

"SNU_res.xlsx"

NCI

Distribution of Log2 Mutant Ratios



Looks reasonable.

Quantile thresholds

Values comparing log2 ratios of mutants to wt will be determined to be of interest if they are beyond these thresholds.

- **quantiles**:

	p=0.025	p=0.975
estimate	-1.845	1.859

- **probs**: *0.025* and *0.975*

Comparison between BAP1 and EV

EV is the reference.

Genes up regulated in BAP1

Ensembl.ID	Gene.Name	NCIEV_V1N_1	NCIBAP1_V1N_1	log2_ratio
ENSG00000170345	FOS	0.3027	22.2	6.196
ENSG00000125740	FOSB	0.6055	13.77	4.507
ENSG00000268223	ARL14EPL	0.07568	1.237	4.031
ENSG00000143867	OSR1	0.07568	1.005	3.732
ENSG00000183018	SPNS2	0.1514	1.934	3.675
ENSG00000138395	CDK15	0.07568	0.8508	3.491
ENSG00000213599	SLX1A-SULT1A3	0.07568	0.8508	3.491
ENSG00000120738	EGR1	1.362	14.54	3.416
ENSG00000174130	TLR6	0.2271	2.32	3.353
ENSG00000189127	ANKRD34B	0.07568	0.7734	3.353

Full up regulated gene list:

"Human_NCI_UpinBAP1.xlsx"

Genes down regulated in BAP1

Ensembl.ID	Gene.Name	NCIEV_V1N_1	NCIBAP1_V1N_1	log2_ratio
ENSG00000279484	KLHL30-AS1	1.135	0.07734	-3.876
ENSG00000226179	LINC00685	1.06	0.07734	-3.776
ENSG00000128482	RNF112	0.9839	0.07734	-3.669
ENSG00000215417	MIR17HG	0.8325	0.07734	-3.428
ENSG00000179909	ZNF154	1.514	0.1547	-3.291

Ensembl.ID	Gene.Name	NCIEV_V1N_1	NCIBAP1_V1N_1	log2_ratio
ENSG00000233266	HMGB1P31	0.7568	0.07734	-3.291
ENSG00000228331	RPL17P43	0.7568	0.07734	-3.291
ENSG00000225573	RPL35P5	0.7568	0.07734	-3.291
ENSG00000255568	BRWD1-AS2	0.6812	0.07734	-3.139
ENSG00000282100	HSP90AB4P	0.6812	0.07734	-3.139

Full up regulated gene list:

"Mouse_B16_DowninMut.xlsx"

Approximate FDR for the lists

Rough FDR for this list is 1. This is because we see fewer significant genes than we expect by chance based on the two mutant comparison. That is, the two mutant lines are less similar than the NCI samples to each other. This could be because there are true differences between the two mutant cell lines, so our null distribution is not really null.

Conclusions

None of the genes are significant after correcting for the number of tests. This is likely because of the low power resulting from the low numbers.

Environment

This analysis was run in the following directory:

```
## [1] "../Data/Abdel_Rahman/BAP1_Experiment_05012020"
```

This analysis was run in the following software environment:

```
## R version 3.4.4 (2018-03-15)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 18.04.4 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblasp-r0.2.20.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8          LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8           LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8       LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8          LC_NAME=en_US.UTF-8
##  [9] LC_ADDRESS=en_US.UTF-8        LC_TELEPHONE=en_US.UTF-8
## [11] LC_MEASUREMENT=en_US.UTF-8    LC_IDENTIFICATION=en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] fitdistrplus_1.0-14 npsurv_0.4-0.1      lsei_1.2-0.1
##  [4] MASS_7.3-51.4       pbapply_1.4-1       tibble_2.1.3
##  [7] data.table_1.12.2   printr_0.1          xlsx_0.6.1
## [10] survplot_0.0.7      survival_2.44-1.1   sjmisc_2.8.1
## [13] papeR_1.0-4         xtable_1.8-4        car_3.0-2
## [16] carData_3.0-2       gdata_2.18.0        tableone_0.10.0
## [19] pander_0.6.3
##
## loaded via a namespace (and not attached):
##  [1] gtools_3.8.1     tidyselect_0.2.5 sjlabelled_1.1.0 xfun_0.9
##  [5] purrr_0.3.3      mitools_2.4      rJava_0.9-9      splines_3.4.4
##  [9] haven_2.2.0      lattice_0.20-38  vctrs_0.2.4      htmltools_0.4.0
## [13] yaml_2.2.0       rlang_0.4.5      pillar_1.4.2     foreign_0.8-72
## [17] glue_1.3.1       DBI_1.0.0        readxl_1.3.1     stringr_1.4.0
## [21] cellranger_1.1.0 zip_2.0.3        evaluate_0.14    knitr_1.24
## [25] rio_0.5.16       forcats_0.4.0    parallel_3.4.4   curl_4.0
## [29] xlsxjars_0.6.1   Rcpp_1.0.2       abind_1.4-5      hms_0.5.3
## [33] digest_0.6.20    stringi_1.4.3    gmodels_2.18.1   openxlsx_4.1.0.1
```

```
## [37] insight_0.4.1    dplyr_0.8.3      survey_3.36      grid_3.4.4
## [41] tools_3.4.4      magrittr_1.5     crayon_1.3.4     pkgconfig_2.0.2
## [45] Matrix_1.2-17    assertthat_0.2.1 rmarkdown_1.15   R6_2.4.0
## [49] compiler_3.4.4
```