

Statistics for data from Fig 2h Caillaud *et al* PLoS Biology

Dan MacLean

13 August 2015

Pre-processing

Marie-Cecille Caillaud (MCC) sent me an Excel file of all the haemocytometry measurements she made - file `raw/MCC-Dan_corrected.xlsx`. The file annotates figures from the same biological replicates as colours which I can't parse programmatically. I therefore added columns to the sheet stating the replicate number. I also removed spaces in column headers and saved the file as `raw/MCC-Dan_corrected_Reps_added.xlsx` and exported the sheet with the data to a csv file `fig_2h_data_manual.csv` which I can operate on programmatically and will use as my input.

Use some Python to get the data file into better shape

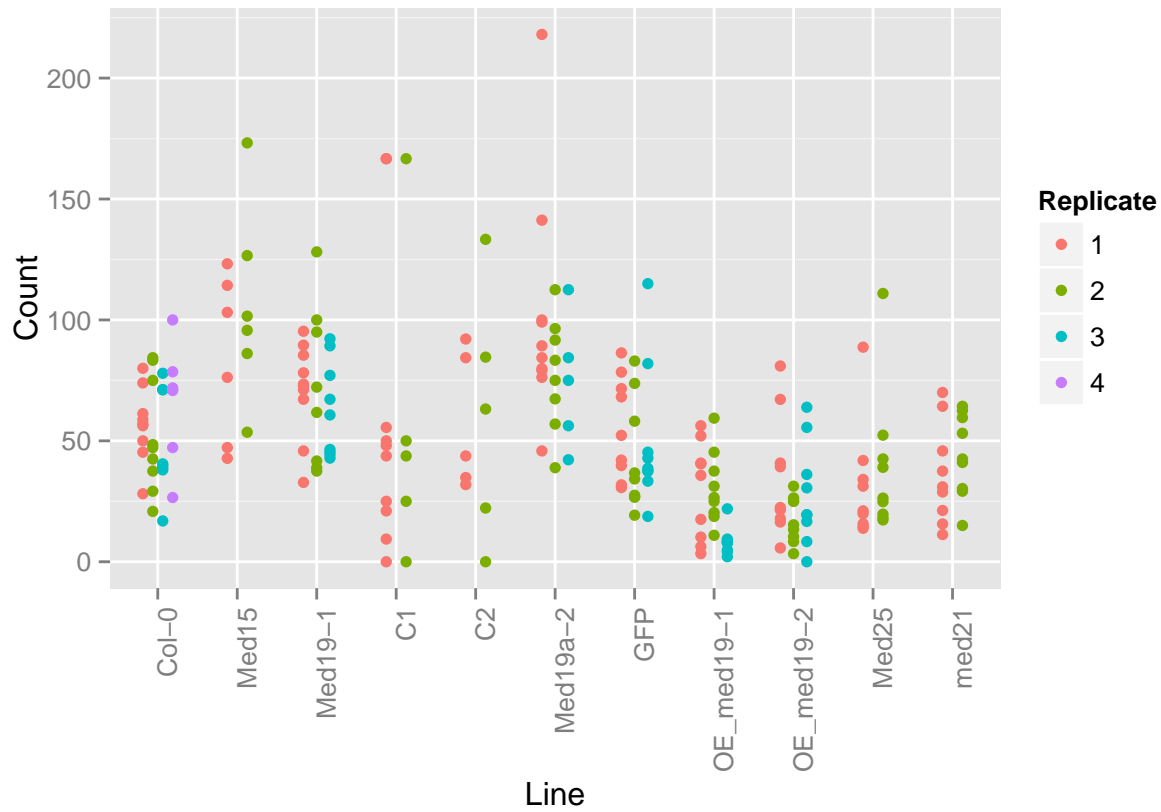
```
header = []
results = []
with open('raw/fig_2h_data_manual.csv', 'r') as file:
    for l in file:
        l = l.rstrip('\r\n')
        a = l.split(',')
        if l.startswith("Rep"):
            header = a
        else:
            for i in range(0, len(header), 2):
                rep, line, count = a[i], header[i+1], a[i+1]
                if rep and line and count: ## if we have no empty values
                    results.append([rep, line, count])

with open('data/reshaped_data.csv', 'w') as outfile:
    outfile.write("Replicate,Line,Count\n")
    for r in results:
        outfile.write(",".join(r) + "\n")
```

Load data, reorder as per Figure 2H and do a straightforward plot

```
library(ggplot2)
data <- read.csv('data/reshaped_data.csv', header=TRUE)
data$Replicate <- as.factor(data$Replicate)
data$Line <- factor(data$Line, c("Col-0", "Med15", "Med19-1", "C1", "C2", "Med19a-2", "GFP", "OE_med19-1", "OE_med19-2"))
basic <- ggplot(data, aes(Line, Count))
scatter <- basic + geom_jitter(aes(colour=Replicate), position = position_dodge(width=0.5)) + theme(axis
```

```
## ymax not defined: adjusting position using y instead
```



data

##	Replicate	Line	Count
## 1	1	Col-0	61.25
## 2	1	Med19-1	78.13
## 3	1	Med15	76.25
## 4	1	Med19a-2	100.00
## 5	1	GFP	30.68
## 6	1	OE_med19-1	6.25
## 7	1	OE_med19-2	16.43
## 8	1	med21	64.29
## 9	1	Med25	14.58
## 10	1	C2	43.75
## 11	1	C1	48.08
## 12	1	Col-0	45.31
## 13	1	Med19-1	32.81
## 14	1	Med15	42.76
## 15	1	Med19a-2	79.86
## 16	1	GFP	31.82
## 17	1	OE_med19-1	52.04
## 18	1	OE_med19-2	40.82
## 19	1	med21	11.25
## 20	1	Med25	19.87
## 21	1	C2	84.38
## 22	1	C1	55.56
## 23	1	Col-0	56.25
## 24	1	Med19-1	95.31
## 25	1	Med15	103.13

## 26	1	Med19a-2	99.11
## 27	1	GFP	78.41
## 28	1	OE_med19-1	40.48
## 29	1	OE_med19-2	39.29
## 30	1	med21	21.20
## 31	1	Med25	15.74
## 32	1	C2	31.91
## 33	1	C1	166.67
## 34	1	Col-0	57.03
## 35	1	Med19-1	71.14
## 36	1	Med15	47.22
## 37	1	Med19a-2	218.06
## 38	1	GFP	39.77
## 39	1	OE_med19-1	56.25
## 40	1	OE_med19-2	22.45
## 41	1	med21	45.83
## 42	1	Med25	34.03
## 43	1	C2	34.82
## 44	1	C1	9.38
## 45	1	Col-0	28.13
## 46	1	Med19-1	72.50
## 47	1	Med15	123.21
## 48	1	Med19a-2	84.38
## 49	1	GFP	86.36
## 50	1	OE_med19-1	35.71
## 51	1	OE_med19-2	67.14
## 52	1	med21	28.85
## 53	1	Med25	13.84
## 54	1	C2	92.11
## 55	1	C1	21.01
## 56	1	Col-0	58.75
## 57	1	Med19-1	85.42
## 58	1	Med15	114.29
## 59	1	Med19a-2	79.17
## 60	1	GFP	42.05
## 61	1	OE_med19-1	10.20
## 62	1	OE_med19-2	5.71
## 63	1	med21	15.63
## 64	1	Med25	31.25
## 65	2	C2	133.33
## 66	1	C1	0.00
## 67	1	Col-0	50.00
## 68	1	Med19-1	73.44
## 69	2	Med15	173.21
## 70	1	Med19a-2	89.29
## 71	1	GFP	71.59
## 72	1	OE_med19-1	40.82
## 73	1	OE_med19-2	80.95
## 74	1	med21	37.50
## 75	1	Med25	41.88
## 76	2	C2	63.16
## 77	1	C1	25.00
## 78	1	Col-0	80.00
## 79	1	Med19-1	67.19

## 80	2	Med15	53.57
## 81	1	Med19a-2	76.25
## 82	1	GFP	68.18
## 83	1	OE_med19-1	17.53
## 84	1	OE_med19-2	21.43
## 85	1	med21	30.98
## 86	1	Med25	21.05
## 87	2	C2	84.62
## 88	1	C1	43.75
## 89	1	Col-0	73.96
## 90	1	Med19-1	89.58
## 91	2	Med15	95.70
## 92	1	Med19a-2	141.25
## 93	1	GFP	52.27
## 94	1	OE_med19-1	3.37
## 95	1	OE_med19-2	17.86
## 96	1	med21	70.00
## 97	1	Med25	88.75
## 98	2	C2	22.22
## 99	1	C1	166.67
## 100	2	Col-0	37.50
## 101	1	Med19-1	45.83
## 102	2	Med15	126.60
## 103	1	Med19a-2	45.83
## 104	2	GFP	58.13
## 105	2	OE_med19-1	26.56
## 106	2	OE_med19-2	10.42
## 107	2	med21	53.13
## 108	2	Med25	18.23
## 109	2	C2	0.00
## 110	1	C1	50.00
## 111	2	Col-0	48.44
## 112	2	Med19-1	61.76
## 113	2	Med15	101.56
## 114	2	Med19a-2	38.89
## 115	2	GFP	27.50
## 116	2	OE_med19-1	37.50
## 117	2	OE_med19-2	3.33
## 118	2	med21	30.26
## 119	2	Med25	24.84
## 120	2	C1	0.00
## 121	2	Col-0	83.33
## 122	2	Med19-1	41.67
## 123	2	Med15	86.11
## 124	2	Med19a-2	96.43
## 125	2	GFP	36.75
## 126	2	OE_med19-1	45.31
## 127	2	OE_med19-2	8.33
## 128	2	med21	42.50
## 129	2	Med25	19.68
## 130	2	C1	25.00
## 131	2	Col-0	84.38
## 132	2	Med19-1	37.50
## 133	2	Med19a-2	112.50

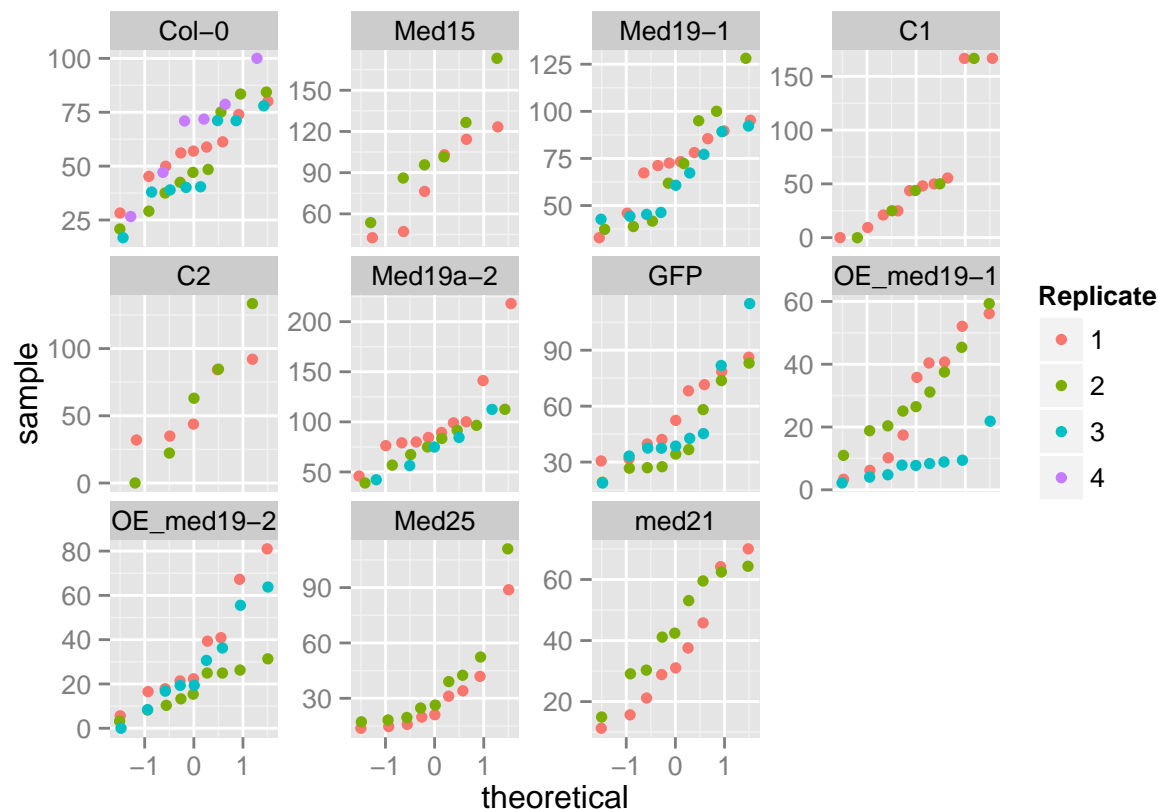
## 134	2	GFP	73.75
## 135	2	OE_med19-1	20.31
## 136	2	OE_med19-2	25.00
## 137	2	med21	64.29
## 138	2	Med25	42.53
## 139	2	C1	43.75
## 140	2	Col-0	42.50
## 141	2	Med19-1	72.22
## 142	2	Med19a-2	67.35
## 143	2	GFP	83.00
## 144	2	OE_med19-1	59.38
## 145	2	OE_med19-2	13.33
## 146	2	med21	62.50
## 147	2	Med25	17.30
## 148	2	C1	166.67
## 149	2	Col-0	20.83
## 150	2	Med19-1	95.00
## 151	2	Med19a-2	75.00
## 152	2	GFP	26.88
## 153	2	OE_med19-1	18.75
## 154	2	OE_med19-2	25.00
## 155	2	med21	15.00
## 156	2	Med25	39.06
## 157	2	C1	50.00
## 158	2	Col-0	75.00
## 159	2	Med19-1	128.13
## 160	2	Med19a-2	83.33
## 161	2	GFP	26.67
## 162	2	OE_med19-1	31.25
## 163	2	OE_med19-2	31.25
## 164	2	med21	29.17
## 165	2	Med25	52.34
## 166	2	Col-0	29.17
## 167	2	Med19-1	100.00
## 168	2	Med19a-2	56.94
## 169	2	GFP	19.25
## 170	2	OE_med19-1	10.94
## 171	2	OE_med19-2	26.39
## 172	2	med21	59.62
## 173	2	Med25	26.32
## 174	2	Col-0	47.22
## 175	2	Med19-1	38.89
## 176	2	Med19a-2	91.67
## 177	2	GFP	34.25
## 178	2	OE_med19-1	25.00
## 179	2	OE_med19-2	15.28
## 180	2	med21	41.07
## 181	2	Med25	110.94
## 182	3	Col-0	16.88
## 183	3	Med19-1	44.32
## 184	3	Med19a-2	75.00
## 185	3	GFP	42.86
## 186	3	OE_med19-1	7.81
## 187	3	OE_med19-2	0.00

## 188	3	Col-0	71.15
## 189	3	Med19-1	89.29
## 190	3	Med19a-2	56.25
## 191	3	GFP	18.75
## 192	3	OE_med19-1	21.88
## 193	3	OE_med19-2	36.11
## 194	3	Col-0	71.15
## 195	3	Med19-1	60.71
## 196	3	Med19a-2	112.50
## 197	3	GFP	33.33
## 198	3	OE_med19-1	7.81
## 199	3	OE_med19-2	30.56
## 200	3	Col-0	38.89
## 201	3	Med19-1	45.31
## 202	3	Med19a-2	84.38
## 203	3	GFP	45.31
## 204	3	OE_med19-1	8.75
## 205	3	OE_med19-2	19.44
## 206	3	Col-0	40.00
## 207	3	Med19-1	67.19
## 208	3	Med19a-2	42.19
## 209	3	GFP	81.94
## 210	3	OE_med19-1	4.17
## 211	3	OE_med19-2	63.89
## 212	3	Col-0	38.00
## 213	3	Med19-1	42.86
## 214	3	GFP	38.64
## 215	3	OE_med19-1	8.33
## 216	3	OE_med19-2	16.67
## 217	3	Col-0	77.94
## 218	3	Med19-1	92.19
## 219	3	GFP	37.50
## 220	3	OE_med19-1	4.86
## 221	3	OE_med19-2	55.56
## 222	3	Col-0	40.48
## 223	3	Med19-1	77.08
## 224	3	GFP	115.00
## 225	3	OE_med19-1	9.38
## 226	3	OE_med19-2	19.44
## 227	4	Col-0	100.00
## 228	3	Med19-1	46.43
## 229	3	GFP	37.50
## 230	3	OE_med19-1	2.08
## 231	3	OE_med19-2	8.33
## 232	4	Col-0	26.56
## 233	4	Col-0	47.22
## 234	4	Col-0	78.57
## 235	4	Col-0	70.83
## 236	4	Col-0	71.88

The data look ok, a few outliers in Med19a-2 and C1 that could affect summary statistics. Let's do some qqplots and see how they lie.

#qnorm is default distribution - we are testing for a normal distribution

```
ggplot(data, aes(sample=Count)) + geom_jitter(stat="qq", aes(colour=Replicate)) + facet_wrap(~ Line, ncol=4)
```



Those outliers could mess up summary statistics, they're off the curve, we have no good reason to ditch them though. I suppose they mean that occasionally the method used (spore counting) throws up some very extreme numbers. Overall these plots are ok, the variation seems normally distributed on the whole.

Let's have a look at summary statistics:

```
library(plyr)
summary <- ddply(data, "Line", summarise, mean=mean(Count), median=median(Count), diff=abs(mean(Count) - median(Count)), std_dev=sd(Count), std_err=se(Count))
summary
```

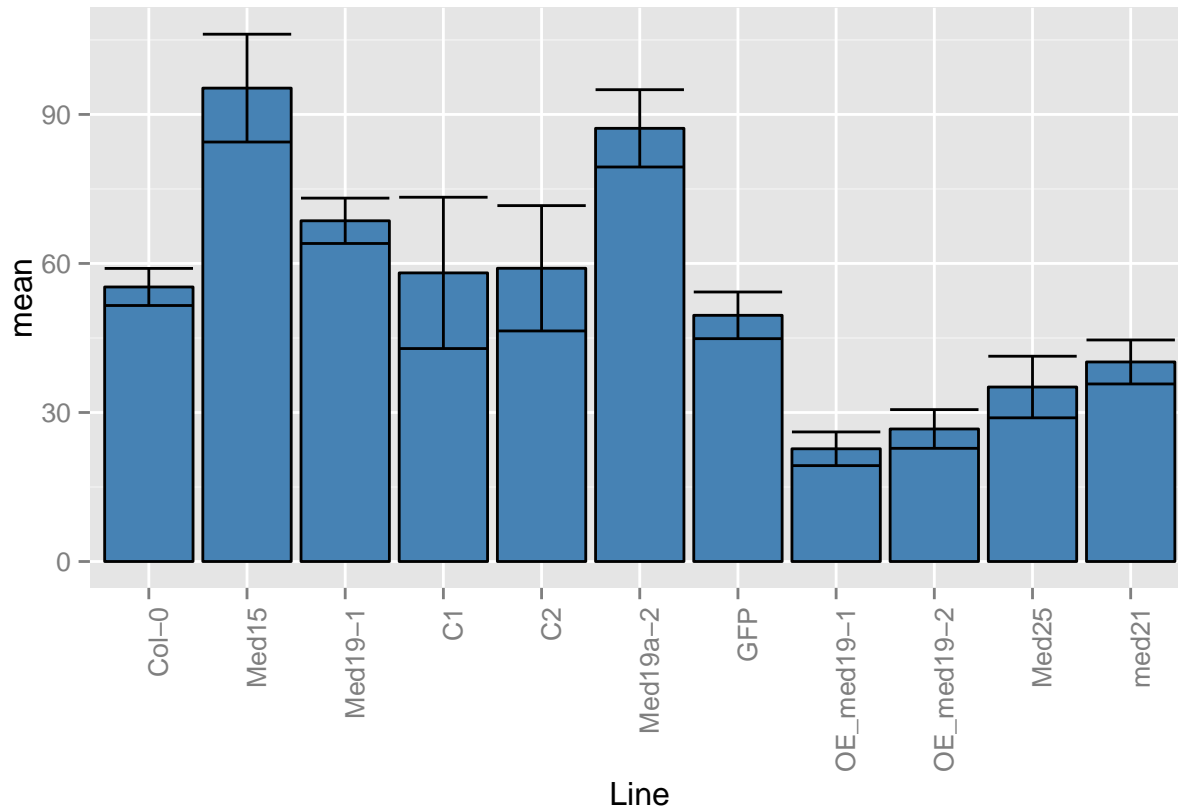
##	Line	mean	median	diff	std_dev	std_err
## 1	Col-0	55.26875	53.125	2.1437500	21.11359	3.732392
## 2	Med15	95.30083	98.630	3.3291667	37.59931	10.853987
## 3	Med19-1	68.58889	71.140	2.5511111	23.73107	4.567046
## 4	C1	58.10267	43.750	14.3526667	58.98184	15.229045
## 5	C2	59.03000	53.455	5.5750000	39.89419	12.615651
## 6	Med19a-2	87.20130	83.330	3.8713043	37.32340	7.782467
## 7	GFP	49.56074	39.770	9.7907407	24.40441	4.696630
## 8	OE_med19-1	22.69333	18.750	3.9433333	17.61271	3.389568
## 9	OE_med19-2	26.68185	21.430	5.2518519	20.25470	3.898019
## 10	Med25	35.12389	25.580	9.5438889	26.30013	6.199001
## 11	med21	40.17056	39.285	0.8855556	18.79839	4.430822

The summary stats seem fine overall, similar SD and SE and not much drift of the median from the mean, the concern again is Med19a-2 and C1 with the high standard deviation and mean dragged up by that couple of points.

Does a bar chart imply a higher effect than we see generally?

Let's make a bar graph with error bars on that first scatter to see how using a standard bar chart might be misleading our thinking.

```
ggplot(summary, aes(x=Line, y=mean)) + geom_bar(position=position_dodge(), stat="identity", fill="steelblue")
```



The barchart is definitely suggesting a higher overall effect than we see from the individual replicates in the scatter plot for Med19a-2 and C1. My conclusion here is that although the mean is calculated correctly, it's just that the mean is a slightly misleading number to boil our data down to in this case. Also that very slight increase in standard error isn't giving us a clue as to that messy single outlier. Taken together the mean and SE plotted like this convince of us a bigger effect in general so the plot style isn't helpful.

Significance Tests

I need to boil down the data to the biological replicates.

```
library(reshape)
```

```
##  
## Attaching package: 'reshape'  
##  
## The following objects are masked from 'package:plyr':  
##  
##   rename, round_any
```



```
bioreps <- cast(data, Line~Replicate, mean)
```

Using Count as value column. Use the value argument to cast to override this choice

```
bioreps <- melt(bioreps)
bioreps
```

##	Line	value	Replicate
## X1	Col-0	56.742222	1
## X1.1	Med15	84.476667	1
## X1.2	Med19-1	71.135000	1
## X1.3	C1	58.612000	1
## X1.4	C2	57.394000	1
## X1.5	Med19a-2	101.320000	1
## X1.6	GFP	55.681111	1
## X1.7	OE_med19-1	29.183333	1
## X1.8	OE_med19-2	34.675556	1
## X1.9	Med25	31.221111	1
## X1.10	med21	36.170000	1
## X2	Col-0	52.041111	2
## X2.1	Med15	106.125000	2
## X2.2	Med19-1	71.896250	2
## X2.3	C1	57.084000	2
## X2.4	C2	60.666000	2
## X2.5	Med19a-2	77.763750	2
## X2.6	GFP	42.908889	2
## X2.7	OE_med19-1	30.555556	2
## X2.8	OE_med19-2	17.592222	2
## X2.9	Med25	39.026667	2
## X2.10	med21	44.171111	2
## X3	Col-0	49.311250	3
## X3.1	Med15	NaN	3
## X3.2	Med19-1	62.820000	3
## X3.3	C1	NaN	3
## X3.4	C2	NaN	3
## X3.5	Med19a-2	74.064000	3
## X3.6	GFP	50.092222	3
## X3.7	OE_med19-1	8.341111	3
## X3.8	OE_med19-2	27.777778	3
## X3.9	Med25	NaN	3
## X3.10	med21	NaN	3
## X4	Col-0	65.843333	4
## X4.1	Med15	NaN	4
## X4.2	Med19-1	NaN	4
## X4.3	C1	NaN	4
## X4.4	C2	NaN	4
## X4.5	Med19a-2	NaN	4
## X4.6	GFP	NaN	4
## X4.7	OE_med19-1	NaN	4
## X4.8	OE_med19-2	NaN	4
## X4.9	Med25	NaN	4
## X4.10	med21	NaN	4

I'll do an ANOVA and Tukey's HSD for multiple comparisons.

```
### ANOVA and Tukey's HSD on all pairwise - though really only interested in VS Col-0 the control
fit <- aov(lm(value ~ Line,data=bioreps))
TukeyHSD(fit)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lm(value ~ Line, data = bioreps))
##
## $Line
##
```

	diff	lwr	upr	p adj
## Med15-Col-0	39.316354	10.505345	68.1273629	0.0033725
## Med19-1-Col-0	12.632604	-12.776317	38.0415256	0.7560958
## C1-Col-0	1.863521	-26.947488	30.6745296	1.0000000
## C2-Col-0	3.045521	-25.765488	31.8565296	0.9999983
## Med19a-2-Col-0	28.398104	2.989183	53.8070256	0.0213201
## GFP-Col-0	-6.423738	-31.832660	18.9851830	0.9961024
## OE_med19-1-Col-0	-33.291146	-58.700067	-7.8822244	0.0050706
## OE_med19-2-Col-0	-29.302627	-54.711549	-3.8937059	0.0163874
## Med25-Col-0	-20.860590	-49.671599	7.9504185	0.2940977
## med21-Col-0	-15.813924	-44.624932	12.9970851	0.6466693
## Med19-1-Med15	-26.683750	-57.053220	3.6857198	0.1149884
## C1-Med15	-37.452833	-70.720921	-4.1847460	0.0200774
## C2-Med15	-36.270833	-69.538921	-3.0027460	0.0260686
## Med19a-2-Med15	-10.918250	-41.287720	19.4512198	0.9539136
## GFP-Med15	-45.740093	-76.109562	-15.3706228	0.0011758
## OE_med19-1-Med15	-72.607500	-102.976970	-42.2380302	0.0000029
## OE_med19-2-Med15	-68.618981	-98.988451	-38.2495117	0.0000066
## Med25-Med15	-60.176944	-93.445032	-26.9088571	0.0001323
## med21-Med15	-55.130278	-88.398365	-21.8621905	0.0003896
## C1-Med19-1	-10.769083	-41.138553	19.6003864	0.9576648
## C2-Med19-1	-9.587083	-39.956553	20.7823864	0.9801671
## Med19a-2-Med19-1	15.765500	-11.397780	42.9287795	0.5771317
## GFP-Med19-1	-19.056343	-46.219622	8.1069370	0.3317637
## OE_med19-1-Med19-1	-45.923750	-73.087030	-18.7604705	0.0003061
## OE_med19-2-Med19-1	-41.935231	-69.098511	-14.7719519	0.0008904
## Med25-Med19-1	-33.493194	-63.862664	-3.1237247	0.0237690
## med21-Med19-1	-28.446528	-58.815998	1.9229420	0.0778077
## C2-C1	1.182000	-32.086087	34.4500873	1.0000000
## Med19a-2-C1	26.534583	-3.834886	56.9040531	0.1187604
## GFP-C1	-8.287259	-38.656729	22.0822105	0.9930153
## OE_med19-1-C1	-35.154667	-65.524136	-4.7851969	0.0158700
## OE_med19-2-C1	-31.166148	-61.535618	-0.7966784	0.0414666
## Med25-C1	-22.724111	-55.992198	10.5439762	0.3649176
## med21-C1	-17.677444	-50.945532	15.5906429	0.6847841
## Med19a-2-C2	25.352583	-5.016886	55.7220531	0.1526233
## GFP-C2	-9.469259	-39.838729	20.9002105	0.9817785
## OE_med19-1-C2	-36.336667	-66.706136	-5.9671969	0.0118798
## OE_med19-2-C2	-32.348148	-62.717618	-1.9786784	0.0313070
## Med25-C2	-23.906111	-57.174198	9.3619762	0.3029590
## med21-C2	-18.859444	-52.127532	14.4086429	0.6069765
## GFP-Med19a-2	-34.821843	-61.985122	-7.6585630	0.0062678

```
## OE_med19-1-Med19a-2    -61.689250  -88.852530  -34.5259705  0.0000061
## OE_med19-2-Med19a-2    -57.700731  -84.864011  -30.5374519  0.0000157
## Med25-Med19a-2         -49.258694  -79.628164  -18.8892247  0.0005027
## med21-Med19a-2         -44.212028  -74.581498  -13.8425580  0.0017075
## OE_med19-1-GFP         -26.867407  -54.030687   0.2958721  0.0540066
## OE_med19-2-GFP         -22.878889  -50.042168   4.2843907  0.1455538
## Med25-GFP              -14.436852  -44.806322  15.9326179  0.7982744
## med21-GFP              -9.390185  -39.759655  20.9792846  0.9828033
## OE_med19-2-OE_med19-1   3.988519  -23.174761  31.1517981  0.9999631
## Med25-OE_med19-1       12.430556  -17.938914  42.8000253  0.9027929
## med21-OE_med19-1       17.477222  -12.892248  47.8466920  0.5879787
## Med25-OE_med19-2        8.442037  -21.927433  38.8115068  0.9919835
## med21-OE_med19-2       13.488704  -16.880766  43.8581735  0.8525396
## med21-Med25             5.046667  -28.221421  38.3147540  0.9999503
```

A long table, but it's showing the overexpressers OE_med19-1 and OE_med19-2 are different from the Col-0 control, as is the one with the noted high outliers Med19a-2 and also Med15.

P-Hacking

Let's see how removing those high (≥ 150) outliers affects the p -values, see if any significance we have is coming from one or two atypical data.

```
under_150 <- data[data$value < 150, ]
bioreps_under150 <- cast(data, Line~Replicate, mean)
```

```
## Using Count as value column. Use the value argument to cast to override this choice
```

```
bioreps_under150 <- melt(bioreps_under150)
fit <- aov(lm(value ~ Line, data=bioreps_under150))
TukeyHSD(fit)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lm(value ~ Line, data = bioreps_under150))
##
## $Line
##              diff          lwr          upr          p adj
## Med15-Col-0    39.316354    10.505345    68.1273629  0.0033725
## Med19-1-Col-0   12.632604   -12.776317    38.0415256  0.7560958
## C1-Col-0        1.863521   -26.947488    30.6745296  1.0000000
## C2-Col-0        3.045521   -25.765488    31.8565296  0.9999983
## Med19a-2-Col-0  28.398104    2.989183    53.8070256  0.0213201
## GFP-Col-0       -6.423738   -31.832660    18.9851830  0.9961024
## OE_med19-1-Col-0 -33.291146   -58.700067   -7.8822244  0.0050706
## OE_med19-2-Col-0 -29.302627   -54.711549   -3.8937059  0.0163874
## Med25-Col-0     -20.860590   -49.671599    7.9504185  0.2940977
## med21-Col-0     -15.813924   -44.624932    12.9970851  0.6466693
## Med19-1-Med15   -26.683750   -57.053220    3.6857198  0.1149884
## C1-Med15        -37.452833   -70.720921   -4.1847460  0.0200774
## C2-Med15        -36.270833   -69.538921   -3.0027460  0.0260686
```

## Med19a-2-Med15	-10.918250	-41.287720	19.4512198	0.9539136
## GFP-Med15	-45.740093	-76.109562	-15.3706228	0.0011758
## OE_med19-1-Med15	-72.607500	-102.976970	-42.2380302	0.0000029
## OE_med19-2-Med15	-68.618981	-98.988451	-38.2495117	0.0000066
## Med25-Med15	-60.176944	-93.445032	-26.9088571	0.0001323
## med21-Med15	-55.130278	-88.398365	-21.8621905	0.0003896
## C1-Med19-1	-10.769083	-41.138553	19.6003864	0.9576648
## C2-Med19-1	-9.587083	-39.956553	20.7823864	0.9801671
## Med19a-2-Med19-1	15.765500	-11.397780	42.9287795	0.5771317
## GFP-Med19-1	-19.056343	-46.219622	8.1069370	0.3317637
## OE_med19-1-Med19-1	-45.923750	-73.087030	-18.7604705	0.0003061
## OE_med19-2-Med19-1	-41.935231	-69.098511	-14.7719519	0.0008904
## Med25-Med19-1	-33.493194	-63.862664	-3.1237247	0.0237690
## med21-Med19-1	-28.446528	-58.815998	1.9229420	0.0778077
## C2-C1	1.182000	-32.086087	34.4500873	1.0000000
## Med19a-2-C1	26.534583	-3.834886	56.9040531	0.1187604
## GFP-C1	-8.287259	-38.656729	22.0822105	0.9930153
## OE_med19-1-C1	-35.154667	-65.524136	-4.7851969	0.0158700
## OE_med19-2-C1	-31.166148	-61.535618	-0.7966784	0.0414666
## Med25-C1	-22.724111	-55.992198	10.5439762	0.3649176
## med21-C1	-17.677444	-50.945532	15.5906429	0.6847841
## Med19a-2-C2	25.352583	-5.016886	55.7220531	0.1526233
## GFP-C2	-9.469259	-39.838729	20.9002105	0.9817785
## OE_med19-1-C2	-36.336667	-66.706136	-5.9671969	0.0118798
## OE_med19-2-C2	-32.348148	-62.717618	-1.9786784	0.0313070
## Med25-C2	-23.906111	-57.174198	9.3619762	0.3029590
## med21-C2	-18.859444	-52.127532	14.4086429	0.6069765
## GFP-Med19a-2	-34.821843	-61.985122	-7.6585630	0.0062678
## OE_med19-1-Med19a-2	-61.689250	-88.852530	-34.5259705	0.0000061
## OE_med19-2-Med19a-2	-57.700731	-84.864011	-30.5374519	0.0000157
## Med25-Med19a-2	-49.258694	-79.628164	-18.8892247	0.0005027
## med21-Med19a-2	-44.212028	-74.581498	-13.8425580	0.0017075
## OE_med19-1-GFP	-26.867407	-54.030687	0.2958721	0.0540066
## OE_med19-2-GFP	-22.878889	-50.042168	4.2843907	0.1455538
## Med25-GFP	-14.436852	-44.806322	15.9326179	0.7982744
## med21-GFP	-9.390185	-39.759655	20.9792846	0.9828033
## OE_med19-2-OE_med19-1	3.988519	-23.174761	31.1517981	0.9999631
## Med25-OE_med19-1	12.430556	-17.938914	42.8000253	0.9027929
## med21-OE_med19-1	17.477222	-12.892248	47.8466920	0.5879787
## Med25-OE_med19-2	8.442037	-21.927433	38.8115068	0.9919835
## med21-OE_med19-2	13.488704	-16.880766	43.8581735	0.8525396
## med21-Med25	5.046667	-28.221421	38.3147540	0.9999503

Looks good! The same Lines come up as significant - the outliers aren't messing with the overall significance result.

More P-Hacking - ditching data originally in Figure 2H!

According to MCC and JJ then the lines of interest are really the med19-1, Med19a-2, OE_med19-1 and OE_med19-2. Let's do the same tests for the restricted set and see if it substantially affects the result.

```
of_interest <- bioreps_under150[bioreps_under150$Line %in% c("Col-0", "Med19-1", "Med19a-2", "OE_med19-1", "OE_med19-2")]
fit <- aov(lm(value ~ Line, data=of_interest))
```

TukeyHSD(fit)

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lm(value ~ Line, data = of_interest))
##
## $Line
##              diff          lwr          upr          p adj
## Med19-1-Col-0    12.632604 -12.105257  37.370466 0.4979780
## Med19a-2-Col-0    28.398104   3.660243  53.135966 0.0228532
## OE_med19-1-Col-0 -33.291146 -58.029007  -8.553284 0.0081105
## OE_med19-2-Col-0 -29.302627 -54.040489  -4.564766 0.0188384
## Med19a-2-Med19-1  15.765500 -10.680386  42.211386 0.3584632
## OE_med19-1-Med19-1 -45.923750 -72.369636 -19.477864 0.0011733
## OE_med19-2-Med19-1 -41.935231 -68.381118 -15.489345 0.0024215
## OE_med19-1-Med19a-2 -61.689250 -88.135136 -35.243364 0.0000889
## OE_med19-2-Med19a-2 -57.700731 -84.146618 -31.254845 0.0001638
## OE_med19-2-OE_med19-1   3.988519 -22.457368  30.434405 0.9869299
```

The result is not substantially different from before, the same lines show up as significantly different, that is Med19a-2, OE_med19-1, OE_med19-2 and Med15 are significantly different from the Col-0 control. Med19-1 is not.

Conclusion

The Med19-2 and Med15 lines get significantly more spores than the Col-0 wild-type and the two over-expressors of Med19 show significantly fewer spores than Col-0. There is no evidence for difference from the wild-type and other lines.

```
## Saving 6.5 x 4.5 in image
## ymax not defined: adjusting position using y instead
```