

# A model to study the biological effects of Epstein-Barr virus in epithelial cells: Bioinformatics summary I

Rene Welch

Department of Statistics  
University of Wisconsin - Madison

November 2nd, 2015

- 1 Alignment statistics
- 2 Gene expression quantification with RSEM
- 3 Differential expression analysis with DESeq
- 4 Comparative analysis

# Alignments statistics

We aligned using the **RSEM's bowtie** default parameters to the **hg19** genome.

## Old data

Treatment	Cell	Replicate	Total Reads	Aligned	Failed	Supressed	Aligned %
CaFBS	NOK	Rep1	20,918,273	9,866,971	10,773,042	278,260	47.17%
CaFBS	NOK	Rep2	19,145,123	7,276,020	11,656,212	212,891	38.0%
CaFBS	NOK	Rep3	14,115,783	6,488,151	7,458,838	168,794	45.96%
MC	NOK	Rep1	25,223,604	15,607,189	9,256,244	360,171	61.88%
MC	NOK	Rep2	23,752,280	15,274,585	8,123,504	354,191	64.31%
MC	NOK	Rep3	25,125,404	13,491,436	11,342,313	291,655	53.7%
No-tr	NOK	Rep1	15,874,751	6,095,284	9,430,327	349,140	38.4%
No-tr	NOK	Rep2	20,729,192	7,489,732	12,908,019	331,441	36.13%
No-tr	NOK	Rep3	39,650,830	15,997,008	23,174,059	479,763	40.34%
No-tr	NOK	Rep4	14,691,745	4,240,920	9,507,712	943,113	28.87%
CaFBS	EBV_NOK	Rep1	46,090,189	25,856,212	19,545,057	688,920	56.1%
CaFBS	EBV_NOK	Rep2	25,832,522	8,947,614	15,786,150	1,098,758	34.64%
CaFBS	EBV_NOK	Rep3	31,548,887	12,327,626	18,283,663	937,598	39.07%
MC	EBV_NOK	Rep1	23,843,595	13,854,296	9,495,500	493,799	58.1%
MC	EBV_NOK	Rep2	22,964,158	11,297,052	11,224,921	442,185	49.19%
MC	EBV_NOK	Rep3	20,190,546	10,793,401	9,003,337	393,808	53.46%
No-tr	EBV_NOK	Rep1	24,738,646	7,319,700	16,692,994	725,952	29.59%
No-tr	EBV_NOK	Rep2	20,133,523	4,317,215	14,318,593	1,497,715	21.44%
No-tr	EBV_NOK	Rep3	44,257,592	16,636,921	25,916,808	1,703,863	37.59%
No-tr	EBV_NOK	Rep4	16,453,801	7,131,873	8,710,491	611,437	43.34%

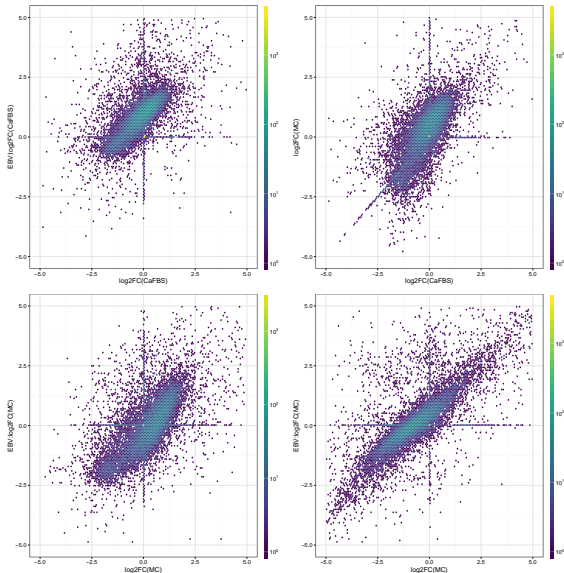
# Alignments statistics

We aligned using the **RSEM's bowtie** default parameters to the **hg19** genome.  
**Scott's data:**

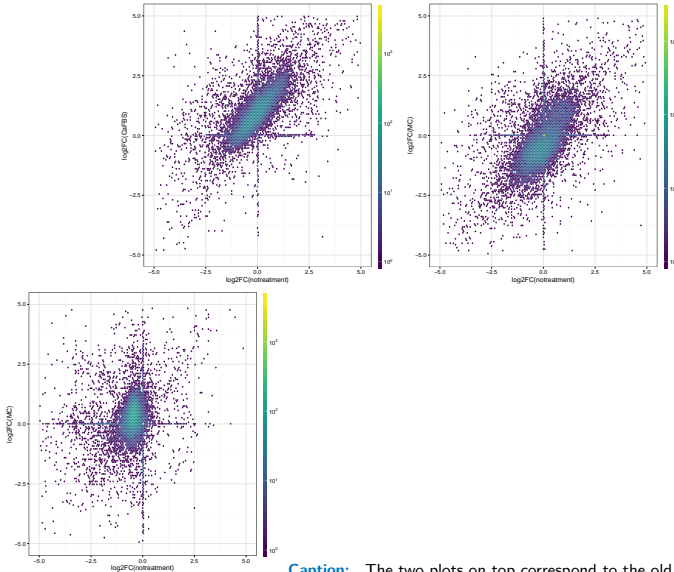
Treatment	Cell	Replicate	Total Reads	Aligned	Failed	Supressed	Aligned %
MC	NOK	Rep1	23,282,428	11,707,297	11,352,280	222,851	50.28%
MC	NOK	Rep2	57,410,179	37,925,563	18,663,347	821,269	66.06%
MC	NOK	Rep3	41,102,488	11,524,597	29,374,338	203,553	28.04%
MC	NOK	Rep4	41,282,292	30,574,767	10,098,455	609,070	74.06%
No-tr	NOK	Rep1	35,298,403	21,367,644	13,535,523	395,236	60.53%
No-tr	NOK	Rep2	36,877,444	28,711,421	7,594,190	571,833	77.86%
No-tr	NOK	Rep3	38,664,441	28,591,454	9,497,438	575,549	73.95%
No-tr	NOK	Rep4	38,382,096	29,780,884	8,002,145	599,067	77.59%
MC	EBV_NOK	Rep1	24,367,026	16,254,296	7,870,336	242,394	66.71%
MC	EBV_NOK	Rep2	33,899,258	19,398,419	14,068,017	432,822	57.22%
MC	EBV_NOK	Rep3	39,523,354	29,660,612	9,226,853	635,889	75.05%
MC	EBV_NOK	Rep4	30,406,432	21,686,918	8,233,310	486,204	71.32%
No-tr	EBV_NOK	Rep1	40,184,017	4,574,064	35,526,438	83,515	11.38%
No-tr	EBV_NOK	Rep2	35,110,520	25,602,610	9,128,544	379,366	72.92%
No-tr	EBV_NOK	Rep3	50,194,500	40,804,450	8,686,319	703,731	81.29%
No-tr	EBV_NOK	Rep4	39,942,772	30,605,482	8,743,082	594,208	76.62%

# Gene expression quantification with RSEM

## (log2 FoldChange counts)



# Gene expression quantification with RSEM (log2 FoldChange counts) [with EBV as treatment]



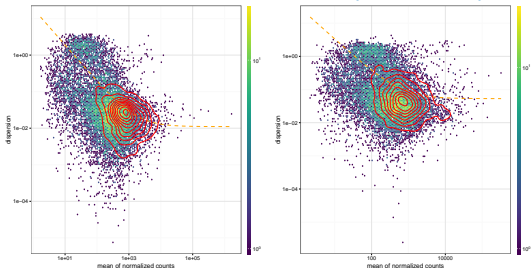
**Caption:** The two plots on top correspond to the old data, and the one below comes from Scott's lab.

# Differential expression analysis with DESeq

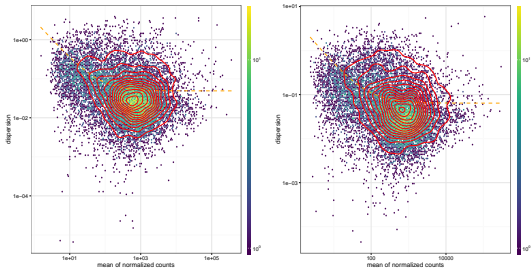
- ▶ **DESeq** model the count data by fitting a NegBin model on the count data.
- ▶ Calculates the dispersion for each gene, and then models the dispersion as a function of the mean gene expression.
- ▶ For this iteration, we are fitting the model for each **Cell.Type** by separate. In further iterations, we are going to fit **Cell.Type** + **Treatment** together.
- ▶ Additionally, need to change the software to **DESeq2** (is a bit more stable).

# Mean gene expression vs. estimated dispersion

## MC vs. no treatment (Old data)



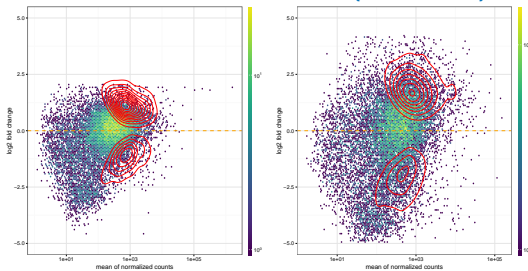
## MC vs. no treatment (Scott's data)



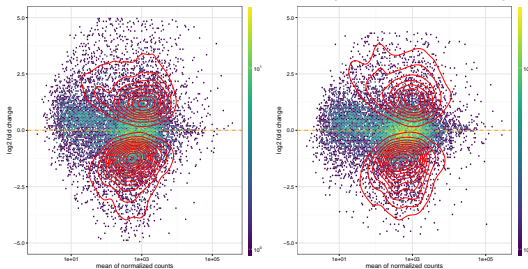


# MA plots

## MC vs. no treatment (Old data)

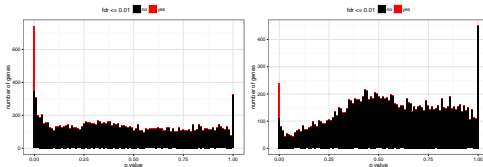


## MC vs. no treatment (Scott's data)

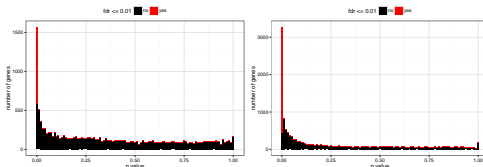


# p.value histograms for diff. expression tests

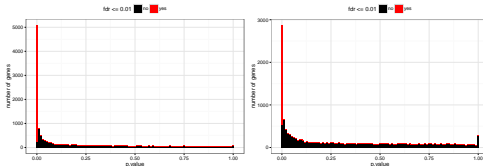
## CaFBS vs. no treatment (Old data)



## MC vs. no treatment (Old data)

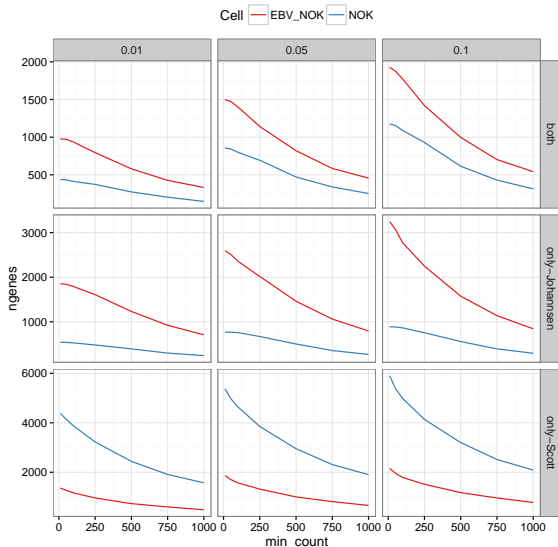


## MC vs. no treatment (Scott's data)



# Comparative analysis

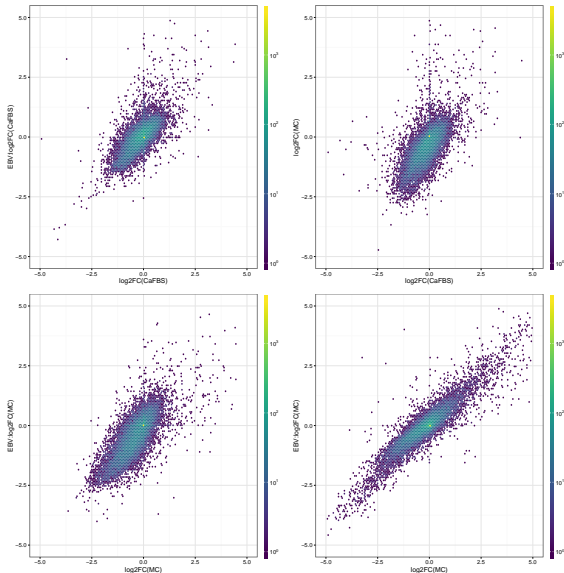
## All genes



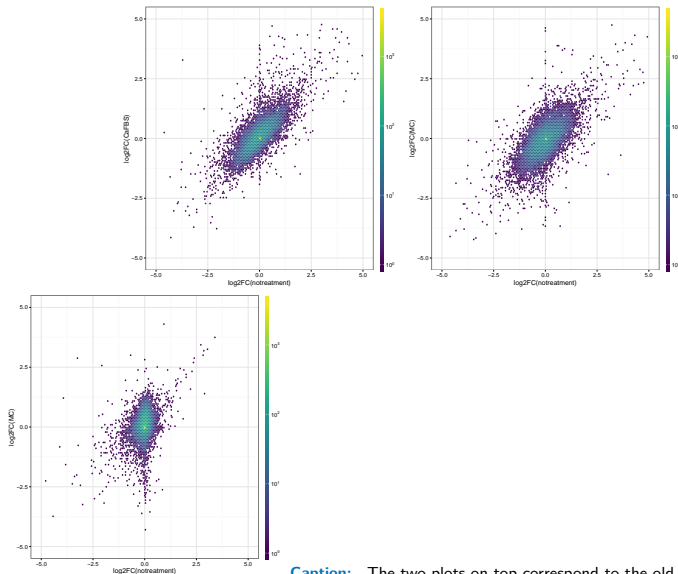
**Additional slides**

# Gene expression quantification with RSEM

## (log2 FoldChange abundance)



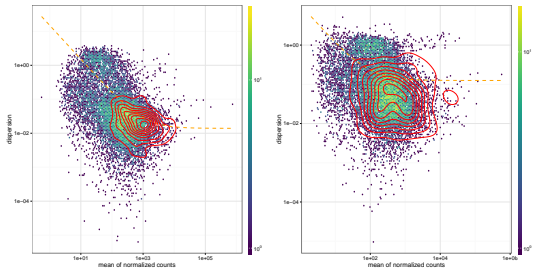
# Gene expression quantification with RSEM (log2 FoldChange abundance) [with EBV as treatment]



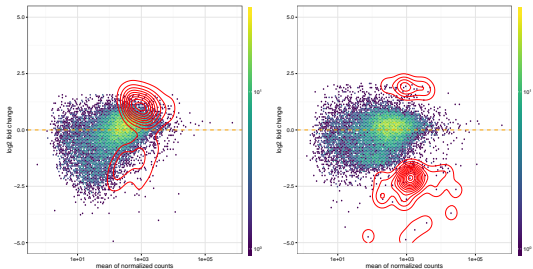
**Caption:** The two plots on top correspond to the old data, and the one below comes from Scott's lab.

# CAFBS diagnostics

## Mean gene expression vs. dispersion

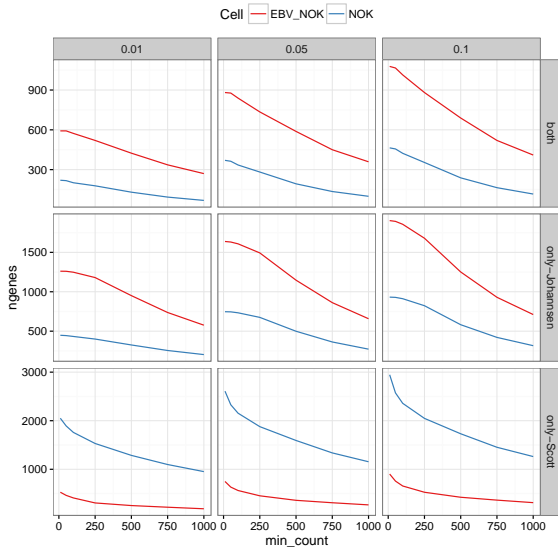


## MA plots



# Comparative analysis

## Upregulated genes





# Comparative analysis

## Downregulated genes

