Title Here

Contents

[Abstract 1](#_Toc500534353)

[1.0 Introduction 2](#_Toc500534354)

[2.0 Methods 2](#_Toc500534355)

[2.1 Data 2](#_Toc500534356)

[2.2 Standard Deviation with LOESS of Moving Window (LMW) 2](#_Toc500534357)

[2.3 Standard Deviation with Smoothing Spline of Absolute Differences (SSAD) 3](#_Toc500534358)

[3](#_Toc500534359)

[2.4 The Agreement Between the Two Methods 3](#_Toc500534360)

[2.4 Test Statistic Distribution and Cut-Offs 3](#_Toc500534361)

[3.0 Results 3](#_Toc500534362)

[4.0 Discussion 4](#_Toc500534363)

[5.0 Acknowledgements 4](#_Toc500534364)

[References 4](#_Toc500534365)

# Abstract

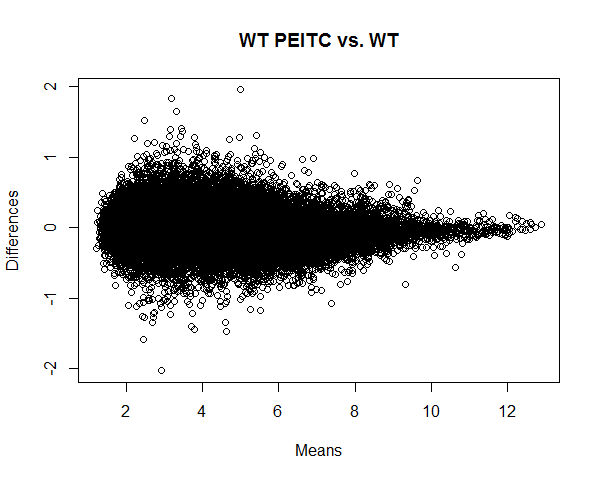
Recent advances in genomics and sharp decreases in gene sequencing price created a flood of data generated by laboratories of all sizes. However, many of these laboratories are operating on strictly constrained budgets and cannot afford large number of samples sequenced for a single project, rerouting resources saved on replication into testing more compounds and models instead. Very often, the results produced by a genome-wide analysis are the first step in screening for genes of interest and are usually validated further by qPCR or pyrosequencing. This may result in a large number of false positives, very quickly diluting the usefulness of such first-pass screening. The purpose of this paper is to describe a statistical approach for evaluation of differential expression of genes in experiment with small number of replicates, including an extreme example of single replicates per treatment. Our method is based on the relationship between the differences and the means of gene expressions. This relationship allows estimation of standard errors by borrowing strength across all genes in the samples. The results indicate that even in experiments with no replicates, useful test statistics can be calculated and applied for filtering of significantly differentially expressed genes. This method does not address the technical variability and assumes that the experimenters’ techniques were reasonably good and no other major technical factor was contributing to heterogeneity of the data. This assumption holds for the data used in this paper that was generated by Dr. T. Kong’s laboratory as we were able to confirm our findings for single replicates using a study with four replicates per group (Yue’s curcumin data reference here).

# 1.0 Introduction

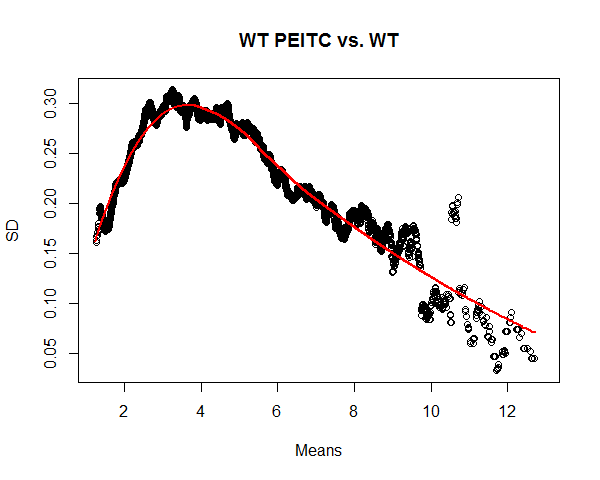
# 2.0 Methods

## 2.1 Data

Multiple data sources were used to test the methods described in this paper. Dataset 1 was generated from a microarray experiment that examined differential expression in genes of wild type (WT) and Sedt7 knock-down (KD) mice treated with phenethyl isothiocyanate (PEICT) compared to controls (Chao’s LNCaP reference and details here).



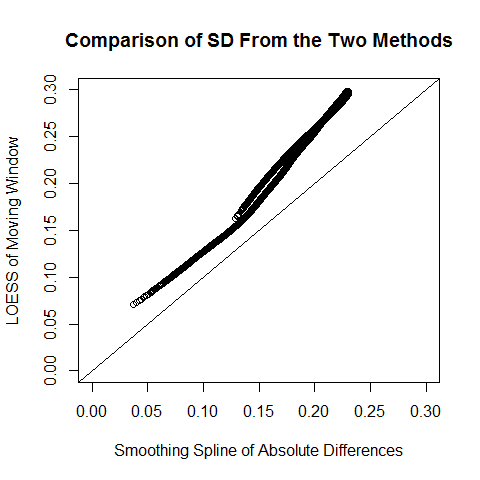
## 2.2 Standard Deviation with LOESS of Moving Window (LMW)



## 2.3 Standard Deviation with Smoothing Spline of Absolute Differences (SSAD)

## 

## 2.4 The Agreement Between the Two Methods

LMW has a higher estimate of variability than SSAD. The differences decrease toward the average SD values.

## 2.4 Test Statistic Distribution and Cut-Offs

NOTE: Prof. Cabrera proposed sampling from controls and estimating SD with SSAD for the controls only. However, there is only 1 sample of control. I cannot randomly sample and reassign expression values as each gene most likely has a range. Otherwise, with random sampling, I completely distort the data structure.

# 3.0 Results

# 4.0 Discussion

# 5.0 Acknowledgements

Everybody in Kong’s lab who is not a coauthor; J&J bioinformaticians that gave me the original idea.

# References

1. Exploration and Analysis of DNA Microarray and Other High-Dimensional Data, 2nd Edition
2. [Question: Microarray Data Without Replicates. Any Hope To Get Something Out Of It?](https://www.biostars.org/p/14130/)
3. [How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC115224/)
4. [The Power of Replicates by Illumina](https://www.illumina.com/Documents/products/technotes/technote_power_replicates.pdf)