

# Transcriptome Analysis of *Mycobacterium leprae*

By: Isabella Fregoso

# What is a transcriptome analysis?



---

Analysis of mRNA transcripts to see the gene expression patterns

*How are the genes expressed?*



---

Quantifies the entire mRNA transcript to tell us about the cellular activity



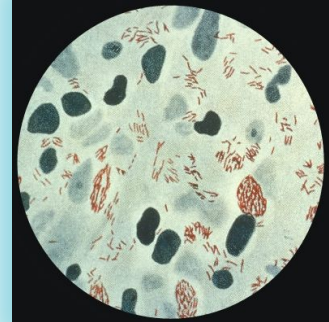
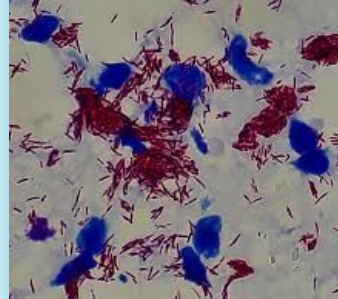
---

RNA-seq used for non coding regions and to identify splice variants

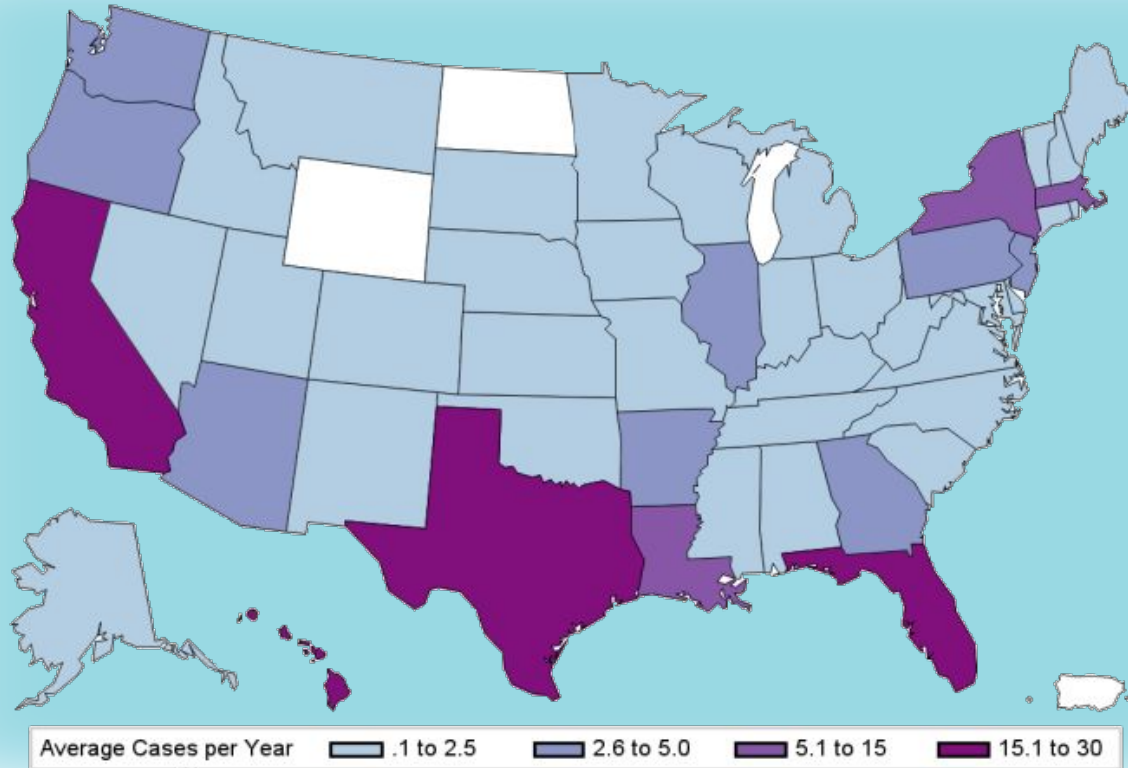
**Splice variant**= different variant of mRNA from same gene, but spliced in a different way

# Introduction

- Acid-fast, rod shaped, gram-positive, obligate intracellular bacterial parasite
- Targets peripheral nerves, epithelial mucosa, eyes, and skin
- Can we map the genes, that code for leprosy?
- Start with gff, fna, and fasta and 2 fastq.gz for each sample



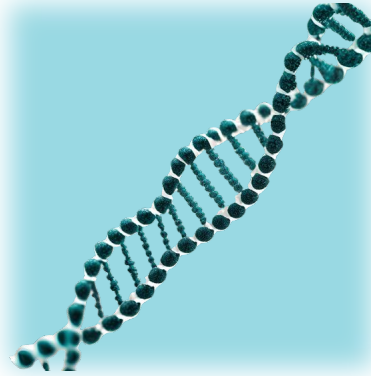
### 10-Year Case Averages by Jurisdiction



Average number of cases of leprosy per year in US (2004–2013) as reported by US HHS

# Methods

- Create index files  
Creates a structured table of contents
- Fastq to sam to bam
- Raw seq data>seq alignment map>  
compressed binary format
- Samtools sort  
Organizes chronologically
- Stringtie assemblies  
RNA seq reads into transcripts

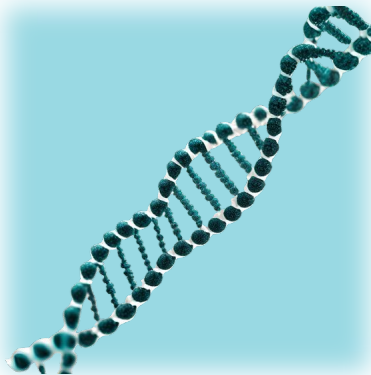


```
.fna  
-----  
.fastq  
.bam  
.sam
```



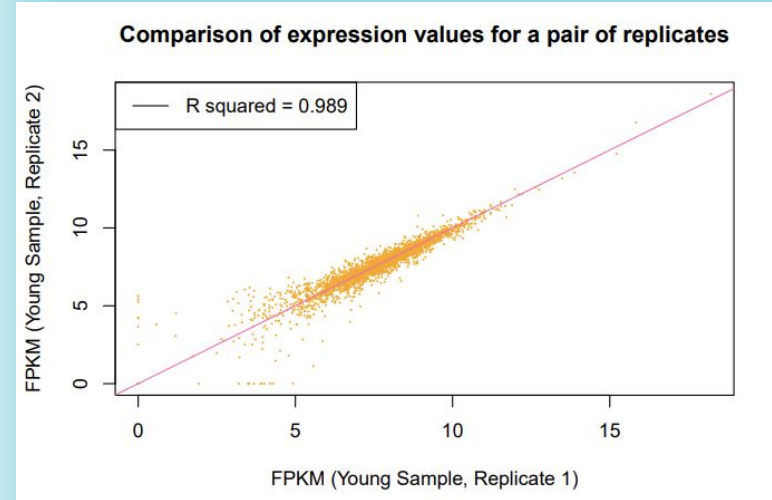
# Methods Cont...

- Stringtie --merge  
Merge transcripts together in single string
- Cat stringtie\_merged.gff  
Counting the transcripts
- Gffcompare -r mleprae.gff  
Compares annotated to known
- Cat merged.status  
Checking the quality



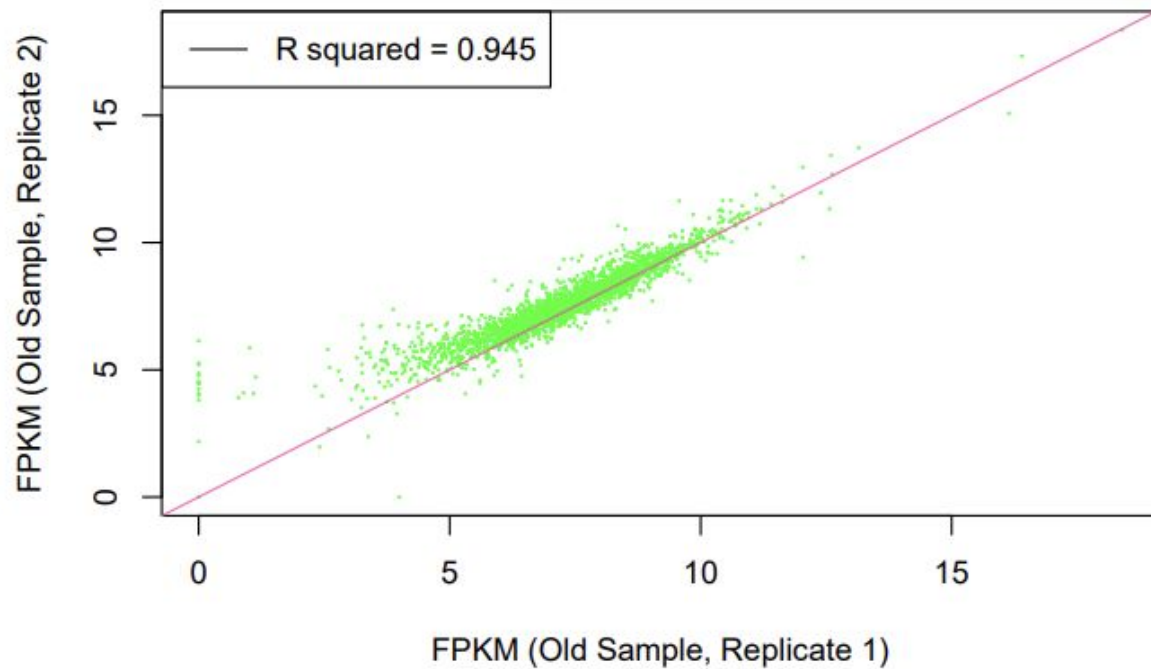
# Results

- $R^2$  closer to one indicates accuracy between predicted and actual expression
- FPKM is Fragment Per Kilobase of transcript per Million mapped reads
- Strong agreement between the replicates; highly consistent and reliable





### Comparison of expression values for a pair of replicates

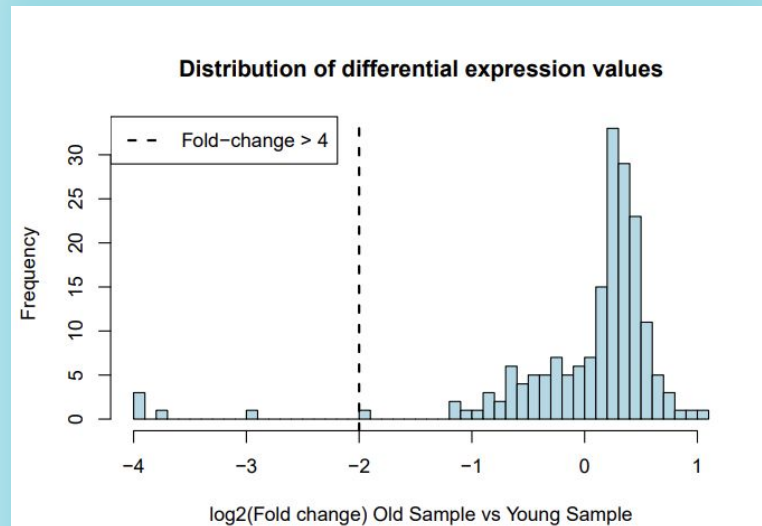




# Results

- This depicts the change in expression between two samples
- Pos fold change means inc in expression, neg fold change means dec in expression

- 
- Black is not significantly differentially expressed
  - Genes above the line are higher expressed in the young, below the line are higher expressed in the old



- Few outliers which indicates significant up/downregulation

## Old vs Young FPKMs

