

cummeRbund: Visualization and Exploration of Cufflinks High-throughput Sequencing Data

Loyal A. Goff, Cole Trapnell

1 April, 2011

Contents

1	Introduction	1
2	CummeRbund Classes	2
2.1	CuffSet Class	2
2.2	CuffData Class	2
2.3	CuffFeatureSet Class	2
2.4	CuffFeature Class	2
3	Reading cuffdiff output	2
3.1	Reading additional annotation files	2
4	Global statistics	2
5	Creating Gene Sets	7
5.1	Geneset level plots	7
6	Individual Genes	7
6.1	Gene-level plots	7
7	Session info	7

1 Introduction

cummeRbund is a visualization package for Cufflinks high-throughput sequencing data. The base class, *cuffSet* is a ‘pointer’ to cufflinks data that are stored out-of-memory in a sqlite database.

2 CummeRbund Classes

- 2.1 CuffSet Class
- 2.2 CuffData Class
- 2.3 CuffFeatureSet Class
- 2.4 CuffFeature Class

3 Reading cuffdiff output

```
> cuff <- readCufflinks(dir = fileDir)
> cuff

CuffSet instance with:
  3 samples
  35695 genes
  102603 isoforms
  54339 TSS
  20584 CDS
```

3.1 Reading additional annotation files

4 Global statistics

Several plotting methods are available that allow for quality-control or global analysis of cufflinks data. For example, to assess the distributions of FPKM scores across samples, you can use the *csDensity* plot (Figure 1).

```
> dens <- csDensity(cuff@genes)
```

Boxplots can be visualized using the *csBoxplot* method (Figure 2).

```
> b <- csBoxplot(cuff@genes)
```

Pairwise comparisons can be made by using *csScatter*. You must specify the sample names to use for the *x* and *y* axes:

```
> s <- csScatter(cuff@genes, "H1_hESC", "Fibroblasts",
+     smooth = T)
```

Volcano plots are also available for the *CuffData* objects. Again, you must specify the comparisons by sample name.

```
> v <- csVolcano(cuff@genes, "H1_hESC", "Fibroblasts")
```

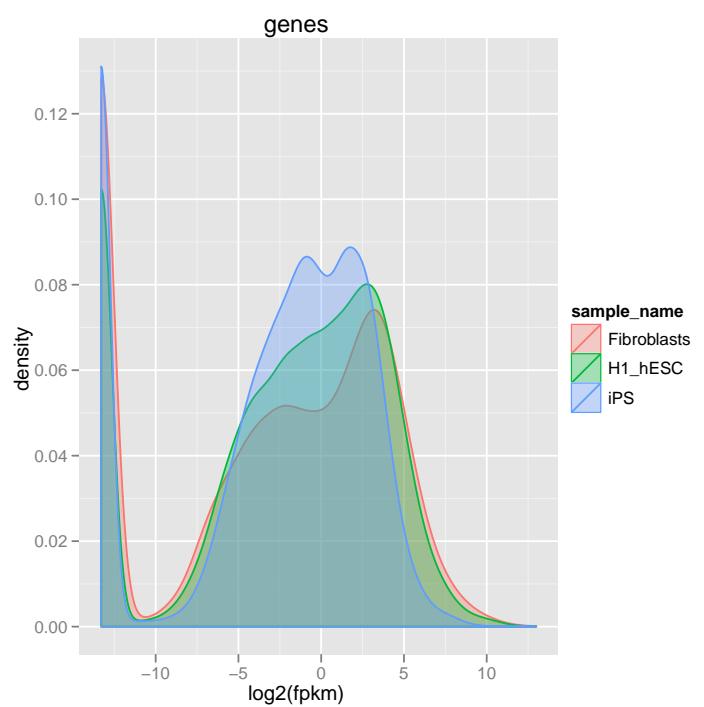


Figure 1: Density plot per sample of cufflinks output FPKM values.

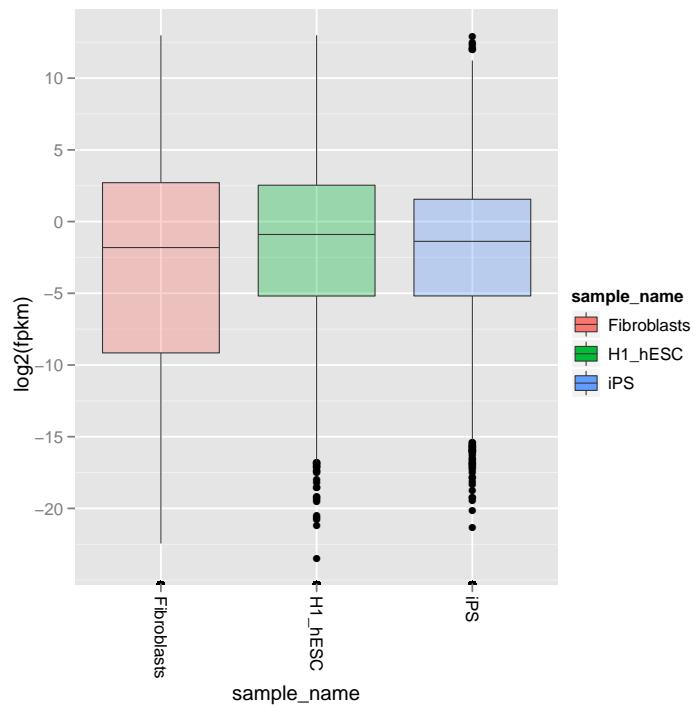


Figure 2: Box plot of FPKM values from cufflinks output.

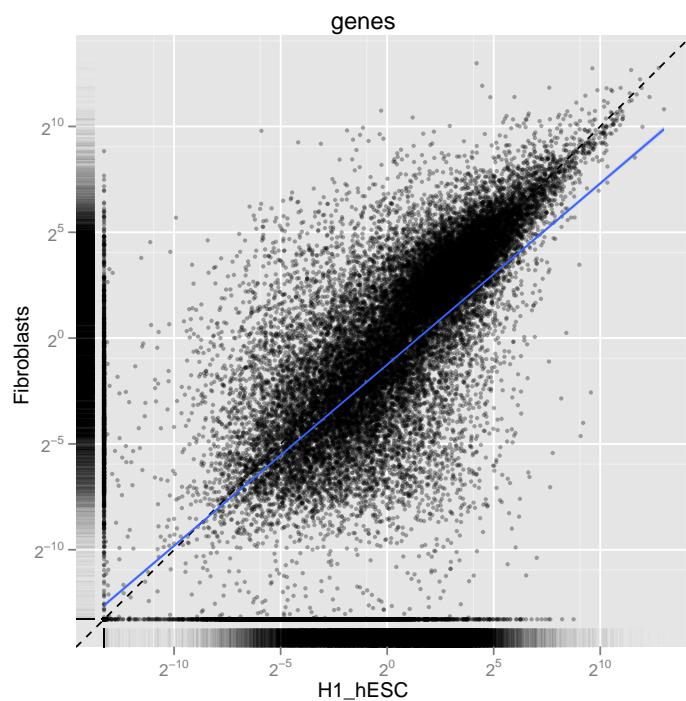


Figure 3: Scatter plot comparing the FPKM values of two samples from cufflinks output.

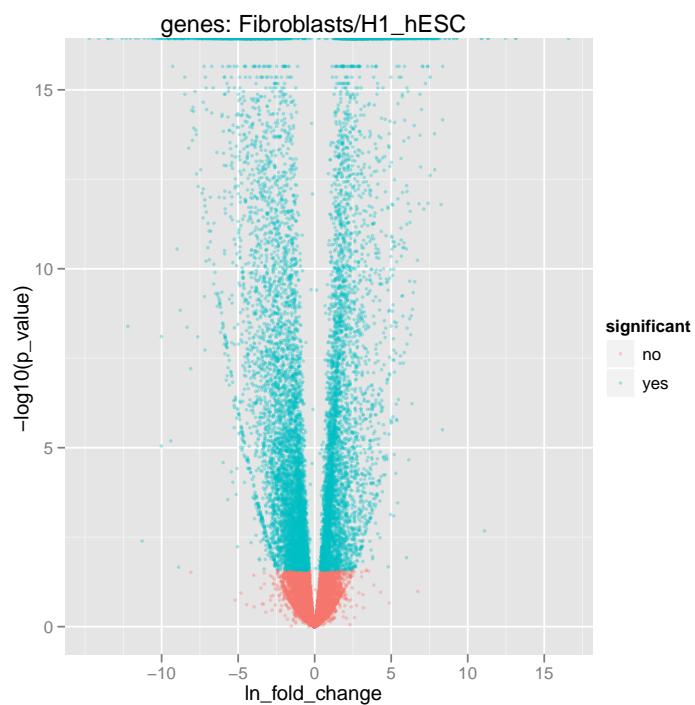


Figure 4: Volcano plot of ln fold change vs significance.

5 Creating Gene Sets

5.1 Geneset level plots

6 Individual Genes

6.1 Gene-level plots

7 Session info

```
> sessionInfo()

R version 2.12.1 (2010-12-16)
Platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit)

locale:
[1] C/en_US.UTF-8/C/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      stats     graphics  grDevices utils      datasets
[7] methods   base

other attached packages:
[1] cummeRbund_0.1.2  ggplot2_0.8.9    proto_0.3-8
[4] reshape_0.8.3     plyr_1.4       RSQLite_0.9-4
[7] DBI_0.2-5

loaded via a namespace (and not attached):
[1] digest_0.4.2 tools_2.12.1
```