

# The trapnell2014myoblasthuman data user's guide

Will Townes (will.townes@gmail.com)

Stephanie C. Hicks (shicks@jimmy.harvard.edu)

Modified: March 17, 2017. Compiled: March 17, 2017

## 1 Overview

---

The trapnell2014myoblasthuman package contains gene expression data from the Trapnell et al. (2014) paper that performed a time-series experiment of bulk (three bulk samples per time point) and single cell RNA-Seq at four time points (hour 0, 24, 48, 72) in differentiated primary human myoblasts ([PMID: 24658644](#)). Metadata and pre-processed data (FPKM) were downloaded from Gene Expression Omnibus ([GSE52529](#)). Batch information was extracted from headers of FASTQ files downloaded from Sequence Read Archive. The data is provided as an ExpressionSet.

The data can be accessed as follows:

```
library(Biobase)
library(trapnell2014myoblasthuman)
```

```
data(trapnell2014myoblasthuman)
```

```
# Get the expression data
eset = exprs(trapnell2014myoblasthuman)
eset[1:5, 1:5]
```

```
##                SRR1032910 SRR1032911 SRR1032912 SRR1032913 SRR1032914
## ENSG00000000003.10  21.984400   7.116630   1.280040   19.6021   43.46180
## ENSG00000000005.5   0.000000   0.000000   0.000000    0.0000   0.00000
## ENSG000000000419.8  40.059700  55.460400  77.580800   77.3450   6.49656
## ENSG000000000457.8   0.937081   0.637528   0.729195   12.4075   0.00000
## ENSG000000000460.12  0.740922   1.405490  57.578500   11.7516   3.93587
```

```
dim(eset)
```

```
## [1] 47192   384
```

```
# Get the pheno data
pd = pData(trapnell2014myoblasthuman)
head(pd)
```

```
##                Run geo_accession      source_name_ch1 description      Model
## SRR1032910 SRR1032910      GSM1268960 Myoblast_Cell TO_CT_A      TO_CT_A01 Illumina HiSeq 2500
## SRR1032911 SRR1032911      GSM1268961 Myoblast_Cell TO_CT_A      TO_CT_A02 Illumina HiSeq 2500
## SRR1032912 SRR1032912      GSM1268962 Myoblast_Cell TO_CT_A      TO_CT_A03 Illumina HiSeq 2500
## SRR1032913 SRR1032913      GSM1268963 Myoblast_Cell TO_CT_A      TO_CT_A04 Illumina HiSeq 2500
## SRR1032914 SRR1032914      GSM1268964 Myoblast_Cell TO_CT_A      TO_CT_A05 Illumina HiSeq 2500
```

```
## SRR1032915 SRR1032915 GSM1268965 Myoblast_Cell T0_CT_A T0_CT_A06 Illumina HiSeq 2500
##          instrument runID      fcID fcLane  hour sampleType          control
## SRR1032910 HWI-ST1233  229 H0L2PADXX      1 hour0          SC control well: FALSE
## SRR1032911 HWI-ST1233  229 H0L2PADXX      1 hour0          SC control well: FALSE
## SRR1032912 HWI-ST1233  229 H0L2PADXX      1 hour0          SC control well: FALSE
## SRR1032913 HWI-ST1233  229 H0L2PADXX      1 hour0          SC control well: FALSE
## SRR1032914 HWI-ST1233  229 H0L2PADXX      1 hour0          SC control well: FALSE
## SRR1032915 HWI-ST1233  229 H0L2PADXX      1 hour0          SC control well: FALSE
##          debris          numcells sampleName
## SRR1032910 debris: FALSE cells in well: 1 T0_CT_A01
## SRR1032911 debris: FALSE cells in well: 1 T0_CT_A02
## SRR1032912 debris: FALSE cells in well: 1 T0_CT_A03
## SRR1032913 debris: FALSE cells in well: 2 T0_CT_A04
## SRR1032914 debris: FALSE cells in well: 1 T0_CT_A05
## SRR1032915 debris: FALSE cells in well: 1 T0_CT_A06
```

The data consists of bulk and single-cell RNA-Seq (SC) at four time points.

```
table(pd$sampleType, pd$hour)

##
##      hour72 hour0 Hour24 Hour48
## SC        84    96    96    96
## bulk        3     3     3     3
```

## 2 References

---

1. Trapnell et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells *Nature Biotechnology* **32** (4): 381 - 386. [PMID: 24658644](#). [PMCID: PMC4122333](#).