

Dynamic Statistical Comparisons

Gao Wang

Matthew Stephens

gaow@uchicago.edu

mstephens@uchicago.edu

University of Chicago <http://github.com/stephenslab>

Simulation data set A.I

Simulation data set A.II

Simulation data set B.I

SVA

myrna

RUV

voom

quasibinom

edgeRglm

DESeq2Glm

limma

ash

jointash

edgeR

DESeq2

score

Give DSC a metric to evaluate DE methods

Finally, construct DSC pipelines by connecting various computational routines

Illustration

Comparison of methods for differential gene expression analysis of RNA sequencing data

Tell DSC we can generate data differently based on what we've done

Tell DSC how benchmark data is generated

```
simulate:
exec: datamaker.R
seed: R(1:50)
params:
  tissue: Adipose-Subcutaneous,
        (Adipose-Subcutaneous, Lung)
  Nsamp: 2, 10, 50
  Ngene: 10000
  breaksample: FALSE, TRUE
  .alias: args = Pack()
return: data, meta = R(data$meta)
```

```
simulate_normalized(simulate):
params:
  voom.normalize: TRUE
```

Tell DSC to consider these confounder control methods

```
correction:
exec: SVA.R,
      RUV.R,
      myrna.R
params:
  data: $data
  .alias: args = Pack()
return: data
```

These are statistical methods tackling the same problem

... and these normalization methods for RNA-Seq data

```
transform:
exec: voom.R, quasibinom.R,
      edgeRglm.R, DESeq2glm.R
params:
  data: $data
  .alias: args = Pack()
return: data
```

... and these differential expression (DE) analysis

```
test:
exec: edgeR.R, DESeq2.R, ash.R,
      jointash.R, limma.R
params:
  data: $data
  .alias: args = Pack()
exec[1,2]:
  glm: TRUE, FALSE
exec[5]:
  robust: TRUE, FALSE
return: output
```

A method may have different flavors (parameters)

Command interface

```
> dsc settings.dsc -j8 --host midway.rcc.uchicago.edu
INFO: DSC script exported to settings.html ...
INFO: Constructing DSC from settings.dsc ...
INFO: Running DSC jobs ...
simulate_1+correction_1+transform_1+test_3+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_1+test_4+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_2+test_3+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_2+test_4+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_3+test_3+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_3+test_4+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_4+test_3+score_1: 100% [=====] 5 20.8/s
simulate_1+correction_1+transform_4+test_4+score_1: 100% [=====] 5 20.8/s
...
simulate_1+test_1: 100% [=====] 2 20.8/s
simulate_1+test_2: 100% [=====] 2 20.8/s
...
DSC: 100% [=====] 116 6.1/s
INFO: Building output database rna_seq.rds ...
INFO: DSC complete!
INFO: Elapsed time 254.829 seconds.
```

```
run: (simulate, simulate_normalized, simulate_partial_null, simulate_thinning) *
      (correction * transform * test[3:4], correction * transform[1] * test[5], test[1:2]) * score
```