Biological interpretation of RNA-seq bioinformatics: a number of pitfalls....

RNA-seq

= big datasets, numerous tools with numerous parameters
  = many different interpretations from the same dataset
Biological interpretation of RNA-seq bioinformatics: a number of pitfalls....

« RNA-seq detects all RNAs »: Is my gene not expressed?
Biological interpretation of RNA-seq bioinformatics: a number of pitfalls....

« RNA-seq detects all RNAs »: Is my gene not expressed?

Starting material:
- Which tissue?
  Laser microdissection vs whole plant
  - Which extraction protocol?
  Small RNAs, ImmunoPrecipitation
  - Which library synthesis protocol?
  polyA selection/ribosomal depletion/size selection
  - Sequencing technology?
    Illumina, PacBio, nanopore, IonTorrent...
  - Sequencing depth?
Biological interpretation of RNA-seq bioinformatics: a number of pitfalls....

« RNA-seq detects all RNAs »: Is my gene not expressed?

Starting material:
- Which tissue?
Laser microdissection vs whole plant
- Which extraction protocol?
Small RNAs, ImmunoPrecipitation
- Which library synthesis protocol?
polyA selection/ribosomal depletion/size selection
- Sequencing technology?
Illumina, PacBio, nanopore, IonTorrent...
- Sequencing depth?
Mapping surprises: the impact of the tools and parameters on the biological interpretation
Splicing analysis in Ath organelles

Plastids

19 introns

Mitochondria

23 introns

Organelles contain small genomes encoding key proteins and RNA necessary for their biology
Mapping surprises: the impact of the tools and parameters on the biological interpretation
Splicing analysis in Ath organelles

Plastids
19 introns

Mitochondria
23 introns

non polyA mRNAs
rRNA depletion
Mapping surprises: the impact of the tools and parameters on the biological interpretation

Splicing analysis in Ath organelles

Plastids

- 19 introns
- non polyA mRNAs
- rRNA depletion
- Splicing analysis
- mapping on genome

Mitochondria

- 23 introns

Mapping surprises: the impact of the tools and parameters on the biological interpretation

Splicing analysis in Ath organelles

Plastids

Mitochondria

Copy of mito genome in Chr2

Mapping on all Chr

19 introns

23 introns

non polyA mRNAs

rRNA depletion

Splicing analysis

mapping on genome
Mapping surprises: the impact of the tools and parameters on the biological interpretation
Splicing analysis in Ath organelles

TopHat2 or STAR?
Genome annotation or not?
Mapping surprises: the impact of the tools and parameters on the biological interpretation
Splicing analysis in Ath organelles

TopHat2 or STAR? Genome annotation or not?
Mapping surprises: the impact of the tools and parameters on the biological interpretation
Splicing analysis in Ath organelles

TopHat2 or STAR? Genome annotation or not?

<table>
<thead>
<tr>
<th></th>
<th>TopHat2</th>
<th>TopHat2 + annot</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C4</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>C5</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C8</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C9</td>
<td>88</td>
<td>95</td>
</tr>
<tr>
<td>C10</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>C11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C12</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>C13</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>C14</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>C15</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>C16</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>C17</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>C18</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C19</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

Nature paper!!!! or biological non sense....
Mapping surprises: the impact of the tools and parameters on the biological interpretation
Splicing analysis in Ath organelles

TopHat2 or STAR? Genome annotation or not?

<table>
<thead>
<tr>
<th></th>
<th>TopHat2</th>
<th>TopHat2 + annot</th>
<th>STAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>11</td>
<td>49</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C4</td>
<td>75</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td>C5</td>
<td>12</td>
<td>15</td>
<td>71</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>C8</td>
<td>0</td>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>C9</td>
<td>88</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>C10</td>
<td>26</td>
<td>27</td>
<td>85</td>
</tr>
<tr>
<td>C11</td>
<td>1</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>C12</td>
<td>62</td>
<td>62</td>
<td>77</td>
</tr>
<tr>
<td>C13</td>
<td>0</td>
<td>27</td>
<td>76</td>
</tr>
<tr>
<td>C14</td>
<td>0</td>
<td>6</td>
<td>88</td>
</tr>
<tr>
<td>C15</td>
<td>0</td>
<td>30</td>
<td>79</td>
</tr>
<tr>
<td>C16</td>
<td>0</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>C17</td>
<td>18</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>C18</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>C19</td>
<td>16</td>
<td>17</td>
<td>66</td>
</tr>
</tbody>
</table>
Mapping surprises: the impact of the tools and parameters on the biological interpretation

Splicing analysis in Ath organelles

TopHat2 or STAR? Genome annotation or not?

<table>
<thead>
<tr>
<th></th>
<th>TopHat2</th>
<th>TopHat2 + annot</th>
<th>STAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>11</td>
<td>49</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C4</td>
<td>75</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td>C5</td>
<td>12</td>
<td>15</td>
<td>71</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>C8</td>
<td>0</td>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>C9</td>
<td>88</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>C10</td>
<td>26</td>
<td>27</td>
<td>85</td>
</tr>
<tr>
<td>C11</td>
<td>1</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>C12</td>
<td>62</td>
<td>62</td>
<td>77</td>
</tr>
<tr>
<td>C13</td>
<td>0</td>
<td>27</td>
<td>76</td>
</tr>
<tr>
<td>C14</td>
<td>0</td>
<td>6</td>
<td>88</td>
</tr>
<tr>
<td>C15</td>
<td>0</td>
<td>30</td>
<td>79</td>
</tr>
<tr>
<td>C16</td>
<td>0</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>C17</td>
<td>18</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>C18</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>C19</td>
<td>16</td>
<td>17</td>
<td>66</td>
</tr>
</tbody>
</table>

Use your biological expertise or borrow it from someone else!!!
Material and methods analysis
One biological interpretation of ..., "differentially expressed"
One biological interpretation of a "differentially expressed" RNA molecule which quantity significantly varies from condition to another.
One biological interpretation of ....

“differentially expressed”

Gene Expression Level: primary spermatocyte vs type B spermatogonia

FDR <= 0.001 AND |log2Ratio| >= 1
- up-regulated genes
- down-regulated genes
- Not DEGs

RNA molecule which quantity significantly varies from condition to another

Genotype, tissue, treatment, physiological state....
One biological interpretation of...

“differentially expressed”

Gene Expression Level: primary spermatocyte vs type B spermatogonia

FDR <= 0.001 AND |log2Ratio| >= 1
- up-regulated genes
- down-regulated genes
- Not DEGs

RNA molecule which quantity significantly varies from condition to another

Statistics and interpretation (thresholds)

Genotype, tissue, treatment, physiological state....
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

Gene Expression Level: primary spermatocyte vs type B spermatogonia

FDR <= 0.001 AND |log2Ratio| >= 1
- up-regulated genes
- down-regulated genes
- Not DEGs

Statistics and interpretation (thresholds)

Genotype, tissue, treatment, physiological state...

Experimental design (study of the parameter of interest)
One biological interpretation of...

"differentially expressed"

Gene Expression Level: primary spermatocyte vs type B spermatogonia

Proportion of the extracted RNA population (no absolute quantification)

RNA molecule which quantity significantly varies from condition to another

Statistics and interpretation (thresholds)

Genotype, tissue, treatment, physiological state....

Experimental design (study of the parameter of interest)
One biological interpretation of differentially expressed RNA molecules which quantity significantly varies from condition to another.

Proportion of the extracted RNA population (no absolute quantification)

Ath 4 days old vs 30 days old leaf

% of reads mapped to:

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>4D</th>
<th>30D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr1</td>
<td>18%</td>
<td>24%</td>
</tr>
<tr>
<td>Chr2</td>
<td>13%</td>
<td>18%</td>
</tr>
<tr>
<td>Chr3</td>
<td>9%</td>
<td>18%</td>
</tr>
<tr>
<td>Chr4</td>
<td>7%</td>
<td>15%</td>
</tr>
<tr>
<td>Chr5</td>
<td>10%</td>
<td>19%</td>
</tr>
<tr>
<td>ChrM</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>ChrC</td>
<td>43%</td>
<td>5%</td>
</tr>
</tbody>
</table>
One biological interpretation of....
“differentially expressed”

Gene Expression Level: primary spermatocyte vs type B spermatogonia

Proportion of the extracted RNA population (no absolute quantification)

Ath 4 days old vs 30 days old leaf

RNA molecule which quantity significantly varies from condition to another

% of reads mapped to:
4D       30D
Chr1     18%       24%
Chr2     13%       18%
Chr3     9%        18%
Chr4     7%        15%
Chr5     10%       19%
ChrM     1%        2%
ChrC     43%       5%

2 fold increase!
One biological interpretation of ....
“differentially expressed”

Gene Expression Level: primary spermatocyte vs type B spermatogonia

Proportion of the extracted RNA population (no absolute quantification)

Ath 4 days old vs 30 days old leaf

% of reads mapped to:

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>4D</th>
<th>30D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr1</td>
<td>18%</td>
<td>24%</td>
</tr>
<tr>
<td>Chr2</td>
<td>13%</td>
<td>18%</td>
</tr>
<tr>
<td>Chr3</td>
<td>9%</td>
<td>18%</td>
</tr>
<tr>
<td>Chr4</td>
<td>7%</td>
<td>15%</td>
</tr>
<tr>
<td>Chr5</td>
<td>10%</td>
<td>19%</td>
</tr>
<tr>
<td>ChrM</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>ChrC</td>
<td>43%</td>
<td>5%</td>
</tr>
</tbody>
</table>

2 fold increase!

NORMALIZATION!!!
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another
One biological interpretation of ....
“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

How?
One biological interpretation of…. “differentially expressed”

RNA molecule which quantity significantly varies from condition to another

How?

Over-expressed RNA
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

How?

Over-expressed RNA \rightarrow Transcription induction
One biological interpretation of “differentially expressed”

RNA molecule which quantity significantly varies from condition to another

How?

Over-expressed RNA → Transcription induction
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

How?

Over-expressed RNA → Transcription induction

Increased stability: miRNA, RNA binding proteins, exosome
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

How?

Over-expressed RNA → Transcription induction

Increased stability: miRNA, RNA binding proteins, exosome
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

How?

Over-expressed RNA → Transcription induction

Increased stability: miRNA, RNA binding proteins, exosome

Additional experiments are required to know "how?"
One biological interpretation of ..... "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

What for?
One biological interpretation of ....
“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context
One biological interpretation of ....
“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context

Experimental design
(study of the parameter of interest)
One biological interpretation of ....
“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

What for?
Differentially expressed RNA = Response to changes of the cellular context

Experimental design
(study of the parameter of interest)

Leaves
/
Floral buds
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context

Experimental design (study of the parameter of interest)

30 fold over-expression of At2g34430 in leaves vs. flower buds
One biological interpretation of ....

“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context

Experimental design
(study of the parameter of interest)

30 fold over-expression of At2g34430 in leaves vs. flower buds

Leaves / Floral buds

\[ \text{At2g34430} = \text{LHCB1.4} : \]

“Light-harvesting chlorophyll-protein complex II subunit B1”
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context

Experimental design (study of the parameter of interest)

30 fold over-expression of At2g34430 in leaves vs. flower buds

Leaves / Floral buds

At2g34430 = LHCBI.4 :
"Light-harvesting chlorophyll-protein complex II subunit B1"

In leaves, there is more photosynthesis than in floral buds so

More LHCBI.4 is required to capture light
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context

Experimental design
(study of the parameter of interest)

30 fold over-expression of \textit{At2g34430} in leaves vs. flower buds

\textit{At2g34430} = \textit{LHCB1.4} :
"Light-harvesting chlorophyll-protein complex II subunit B1"

In leaves, there is more \textit{LHCB1.4} than in floral buds
so
There is more photosynthesis in leaves than in floral buds
One biological interpretation of ... ‘differentially expressed’

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context
One biological interpretation of "differentially expressed"

RNA molecule which quantity significantly varies from condition to another

What for?

Differentially expressed RNA = Response to changes of the cellular context
Differentially expressed RNA = Response to changes of the cellular context

RNA molecule which quantity significantly varies from condition to another

One biological interpretation of .... “differentially expressed”

What for?

Differentially expressed RNA = Response to changes of the cellular context
One biological interpretation of ....
“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

What for?
Differentially expressed RNA = Response to changes of the cellular context

Cadmium treatment in Ath and translational regulation
One biological interpretation of ....

“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

Biomarker?
One biological interpretation of ....
“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

Biomarker?

Differentially expressed RNA  = Specific (?) response to this change of the cellular context
One biological interpretation of ....

“differentially expressed”

RNA molecule which quantity significantly varies from condition to another

Biomarker?

Differentially expressed RNA = Specific (?) response to this change of the cellular context

Expression profile
Biomarker?

Differentially expressed RNA = Specific (?) response to this change of the cellular context

RNA molecule which quantity significantly varies from condition to another

One biological interpretation of .... “differentially expressed”

Expression profile

A single RNA is never specific

Combination of transcripts (multivariate analysis)
One biological interpretation of "differentially expressed"

- RNA involved in the response to the change of the cellular context
- Candidate RNA
- Direct or indirect involvement? Cause or consequence? How?
- One RNA vs several RNAs
- Additional experiments to be done...
One biological interpretation of ....
“differentially expressed”

Differentially accumulated RNA
One biological interpretation of ....

“differentially expressed”

Differentially accumulated RNA

RNA involved in the response to the change of the cellular context
One biological interpretation of ....
“differentially expressed”

Differentially accumulated RNA

RNA involved in the response to the change of the cellular context
One biological interpretation of ....
“differentially expressed”

Differentially accumulated RNA

RNA involved in the response to the change of the cellular context

Direct or indirect involvement?
Cause or consequence?
How?
One biological interpretation of "differentially expressed"

Differentially accumulated RNA

RNA involved in the response to the change of the cellular context

Direct or indirect involvement? Cause or consequence? How?

Additional experiments to be done...
One biological interpretation of “differentially expressed”

Differentially accumulated RNA

RNA involved in the response to the change of the cellular context

Candidate RNA

Direct or indirect involvement?
Cause or consequence?
How?

Additional experiments to be done...
One biological interpretation of "differentially expressed"

Differentially accumulated RNA

RNA involved in the response to the change of the cellular context

Candidate RNA

Direct or indirect involvement?  
Cause or consequence?  
How?

One RNA vs several RNAs

Additional experiments to be done...
One biological interpretation of “co-expressed”
One biological interpretation of “co-expressed”

Genes which expression vary similarly
One biological interpretation of "co-expressed"

Genes which expression vary similarly

Statistics and interpretation
One biological interpretation of “co-expressed”

Genes which expression vary similarly

Ingredients of co-expression

≥ 2 conditions
Variance inter condition (between)
≥ 3 genes

Statistics and interpretation
One biological interpretation of "co-expressed"

Ingredients of co-expression

- $\geq 2$ conditions
- Variance inter condition (between)
- $\geq 3$ genes

Expression profiles

Genes which expression vary similarly

Statistics and interpretation
One biological interpretation of "co-expressed"

Ingredients of co-expression:
- \( \geq 2 \) conditions
- Variance inter condition (between)
- \( \geq 3 \) genes

Genes which expression vary similarly

Statistics and interpretation

Expression profiles

Which genes are more likely grouped together
One biological interpretation of

“co-expressed”

Genes which expression vary similarly

What for?

Co-expressed genes= genes involved in the same responses to the same changes of the cellular context?
= genes involved in the same biological processes?
One biological interpretation of “co-expressed” Genes which expression vary similarly

What for?
> Co-expressed genes → functional partners?
One biological interpretation of “co-expressed”

Genes which expression vary similarly

What for?
>Co-expressed genes → functional partners?

Every protein works with partners
Co-expressed genes → functional partners?

Every protein works with partners

LHCBI.4

Photosystem II
(dimeric)
One biological interpretation of “co-expressed”

Genes which expression vary similarly

What for?

> Co-expressed genes → functional partners?

Every protein works with partners

![Nitrate assimilation pathway diagram]

Photosystem II

$d$imeric
One biological interpretation of “co-expressed”

Genes which expression vary similarly

What for?

>Co-expressed genes → functional partners?
One biological interpretation of “co-expressed”

Genes which expression vary similarly

What for?

> Co-expressed genes → functional partners?

Strong similarity (threshold?)
Numerous conditions (how many?)

Co-expression depends on the tested conditions
One biological interpretation of “co-expressed”

Genes which expression vary similarly

What for?

>Co-expressed genes → functional partners?

Strong similarity (threshold?)
Numerous conditions (how many?)
One biological interpretation of “co-expressed”

Genes which expression vary similarly

How?

Co-expressed genes = under the same regulations?
One biological interpretation of "co-expressed"

Genes which expression vary similarly

How?
Co-expressed genes = under the same regulations?
One biological interpretation of “co-expressed”

Genes which expression vary similarly

How?

Co-expressed genes= under the same regulations?
One biological interpretation of “co-expressed”

Genes which expression vary similarly

How?

Co-expressed genes = under the same regulations?
One biological interpretation of “co-expressed”

Co-expressed genes

Genes potentially involved in the same biological processes given the studied context
One biological interpretation of “co-expressed”

Co-expressed genes

Genes potentially involved in the same biological processes given the studied context
One biological interpretation of “co-expressed”

Co-expressed genes

Genes potentially involved in the same biological processes given the studied context

Putative functional annotation
One biological interpretation of “co-expressed”

Co-expressed genes

Genes potentially involved in the same biological processes given the studied context

Putative functional annotation

Additional experiments to be done...