Ambient temperature regulated flowering time

Applications of RNAseq

RNA-seq course: “The power of RNA-seq”
June 7th, 2013; Richard Immink
Overview

- Introduction: Biological research question/hypothesis
- Experimental set-up and preparations
- Data analysis
- Data confirmation
- Data interpretation: Back to the biological research question
Flowering time: When to flower?
Environmental factors influencing flowering time
Effects of ambient temperature fluctuations
The ambient temperature pathway

For the majority of Regulators no temperature-dependent differential expression!?
Hypothesis

Alternative splicing (AS) is a key mechanism in ambient temperature regulation of flowering time

What is the genome-wide effect of temperature fluctuations on expression and splicing?
Experimental set-up: How to detect and quantify AS events?

**AS characteristics:**
- Events can be rear
- Differences can be limited (few nt only)

**Plant species related questions:**
- Genome sequence available?
- Arrays available?
Tiling array or RNA-seq?

Tiling arrays
mRNA
First strand cDNA synthesis
Second strand cDNA synthesis
Fragmentation using DNase I
Labelling of termini
Hybridization to tiling array
Correlation of expression levels

Spearman’s correlation = 0.90
Exon Boundary detection

RNA-Seq

Tiling Array

Agarwal et al. BMC Genomics 2010, 11:383
http://www.biomedcentral.com/1471-2164/11/383
## RNA-seq!................. But, which method?

<table>
<thead>
<tr>
<th>Method</th>
<th>Seq length (nt)</th>
<th>Read quality</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Illumina Hiseq</strong></td>
<td>fixed 50/100</td>
<td>+++</td>
<td>275 Gb/flowcell</td>
</tr>
<tr>
<td><strong>SOLiD 5500</strong></td>
<td>fixed 75</td>
<td>++++</td>
<td>120 Gb/flowcell</td>
</tr>
<tr>
<td><strong>454/Roche</strong></td>
<td>range av~750</td>
<td>++</td>
<td>1 Mb</td>
</tr>
<tr>
<td><strong>Ion Torrent</strong></td>
<td>range av~170</td>
<td>+</td>
<td>70 Mb/chip</td>
</tr>
<tr>
<td><strong>Pac Bio</strong></td>
<td>range ~3,5 kb</td>
<td>+/-</td>
<td>300.000/run</td>
</tr>
</tbody>
</table>
Illumina Hiseq

Further considerations:

- Sample multiplexing? How many?
- 50 nt or 100 nt run?
- Paired-end or Single read?
RNaseq to search for temperature affected alternative splicing (AS)

Paired-end 100nt Illumina Hiseq runs in duplo, 200-300 bp length fragments, (multiplexing of 4 samples/lane)

Comparison of transcriptome of 23-23 °C control vs. 23-27 °C switch and vs 23-16 °C switch (after 24 hrs!)

Identify genes that are differentially alternatively spliced (and/or differentially expressed).
Sample preparation (Elio Schijlen)

- Delivery of 2-10 μg total RNA
Data analysis (Edouard Severing)

- Millions of reads
  - Demultiplexing
    - Quality control
      - FastQC
    - Raw Fastq
      - Bowtie
      - BWA
      - TopHat
        - DESeq
          - Median count
        - edgeR
          - TMM
        - Summarization
          - e.g., HTSeq
        - baySeq
          - Empirical bayesian
        - MiSO
          - R/FPKM
        - Cufflinks
          - R/FPKM
        - DEXSeq
          - Median count
        - Normalization
          - R/FPKM, median count, TMM

Mapping to the genome

- Transcriptome assembly,
- quantification of events,
- identification of differentially regulated events
Mapping to the genome: Tophat 2.0

- **Settings:**
  - Insert size: (~220; sd ~20; between two adapters)
  - No discordant alignments
    - Alignments have fulfilled the expectations given the insert size and map orientation
    - Alignments have to be paired (no mix mode)
    - Reads aligned to unique regions of the genome are considered
  - Min – max intron size: 50-11000.
Identification of differential expression/splicing: Cufflinks

- Splice-align reads to the genome

- Build a graph representing alternative splicing events

**Settings:**

- Only use of uniquely mapped reads
- Existing annotation is provided to Cufflinks
- Min – max intron size: 50-11000
Results: reproducibility (duplo)
Results

- Total amount of differentially alternatively spliced genes
  - $23^\circ C \rightarrow 27^\circ C$: 1107 genes
  - $23^\circ C \rightarrow 16^\circ C$: 897 genes
  - Reciprocal: 135 genes (15%)

- Differentially alternatively spliced flowering genes
  - $23^\circ C \rightarrow 27^\circ C$: 17 genes
  - $23^\circ C \rightarrow 16^\circ C$: 15 genes
  - Overlap: 7 genes
  - Reciprocal: 4 genes
    (including $FLM$)
RNA-seq results: *FLM* alternative splicing
(mutually exclusive exon)
RNAseq results: *FLM* alternative splicing
(mutually exclusive exon)
Alternative splicing in MAF3
(alternative 3’ splice site)
Alternative splicing in MAF3

(alternative 3’ splice site)
Problem I: Strange events!

MAF2 and MAF3 joined as single gene

At5G65050 = MAF2
At5G65060 = MAF3

- MAF2 and MAF3 joined as single gene
Solution

- Divide genes in variable and constitutive blocks (excluding terminal regions).
- For each variable region we group the isoforms based on their local composition.
Problem II: Mixed reads
Primer design!

- Blast primers against RNA-seq data!

- Note: Analyse based on isoform ratio’s and not absolute values! (semi-quantitative)

(in case to complex: amplify start till end → clone mix of isoforms → sequence individual clones)
Data interpretation: back to our hypothesis

Alternative splicing (AS) is a key mechanism in ambient temperature regulation of flowering time
The *FLM* case

16°C

27°C

- Both *FLMβ* and *FLMδ* are predicted not to be targetted by NMD
- Both *FLMβ* and *FLMδ* transcripts are translated (hence a protein is produced)
Functional analysis *FLM* isoforms

*35S::FLM*\(^\beta\) = repressor of flowering

*35S::FLM*\(^\delta\) = activator of flowering
Conclusion: the ambient temperature dependent flowering time control can be explained by FLM AS!
Acknowledgements

PRI-Bioscience-PDS
Leonie Verhage
Gerco Angenent

MPI Tuebingen
David Posé
Markus Schmid

WUR-Bioinformatics
Edouard Severing
Elio Schijlen

Horizon Breakthrough

NWO
Netherlands Organisation for Scientific Research

Plant Research International Wageningen