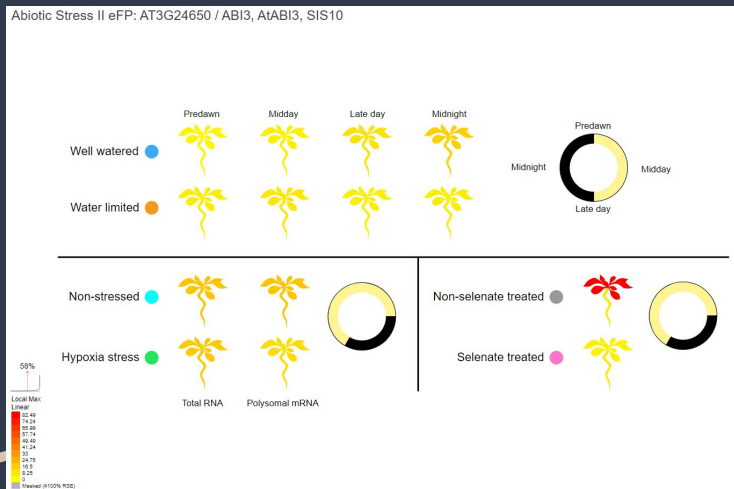


# BCB430 Project

Bruno Pereira,  
1002407086

# Background

Abiotic Stress II eFP: AT3G24650 / ABI3, AtABI3, SIS10



## ePlant eFP Viewer

- Displays expression data in a visual format, combining SVG, XML and RPKM data

```
<?xml version="1.0" encoding="utf-8"?>
<!-- Generator: Adobe Illustrator 16.0.0, SVG Export Plug-In . SVG Version: 6.00 Build 0) -->
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN" "http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">
<svg version="1.1" xmlns="http://www.w3.org/2000/svg" xmlns:xlink="http://www.w3.org/1999/xlink" x="0px" y="0px" width="300px"
height="300px" viewBox="0 0 300 300" enable-background="new 0 0 300 300" xml:space="preserve">
<g id="shapes">
<g id="receptacle">
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-0.874,15.441,0.292,20.686...
</g>
<g id="flower">
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-10.489-1.747-11.946-4.879...
</g>
<g id="outline">
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-0.874,15.441,0.292,20.686...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-10.489-1.747-11.946-4.879...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M208.378,58.782...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M194.376,51.489...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M130.121,161.98...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M91.138,99.985...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M61.564,66.537c-0.185,0.035-0.299,0-0.299,0...
</g>
</svg>
```

```
<?xml version="1.0" encoding="utf-8"?>
<!-- Generator: Adobe Illustrator 16.0.0, SVG Export Plug-In . SVG Version: 6.00 Build 0) -->
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN" "http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">
<svg version="1.1" xmlns="http://www.w3.org/2000/svg" xmlns:xlink="http://www.w3.org/1999/xlink" x="0px" y="0px" width="300px"
height="300px" viewBox="0 0 300 300" enable-background="new 0 0 300 300" xml:space="preserve">
<g id="shapes">
<g id="receptacle">
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-0.874,15.441,0.292,20.686...
</g>
<g id="flower">
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-10.489-1.747-11.946-4.879...
</g>
<g id="outline">
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-0.874,15.441,0.292,20.686...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M120.738,212.084c0,0-10.489-1.747-11.946-4.879...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M208.378,58.782...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M194.376,51.489...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M130.121,161.98...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M91.138,99.985...
<path fill="none" stroke="#000000" stroke-width="0.5" stroke-miterlimit="10" d="M61.564,66.537c-0.185,0.035-0.299,0-0.299,0...
</g>
</svg>
```

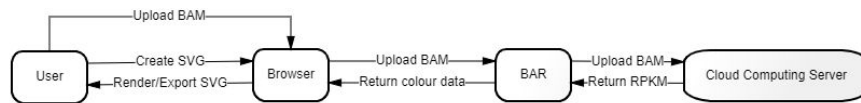
# Proposed project

Web app to allow users to easily create eFP views

Requires:

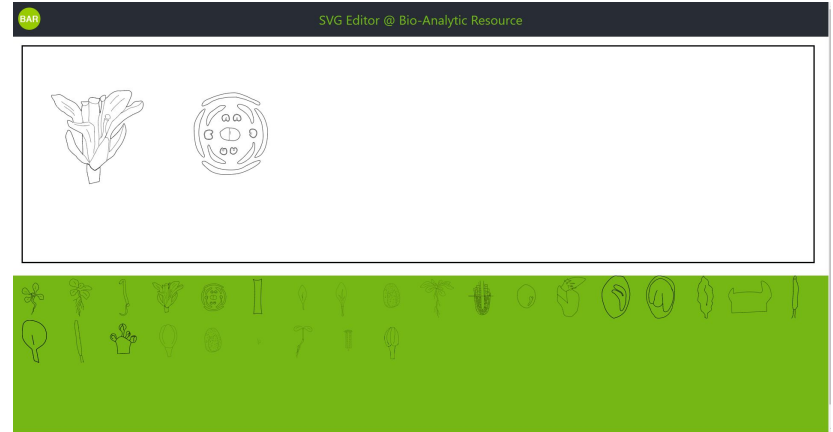
Front-end interface

Back-end BAM file summarization



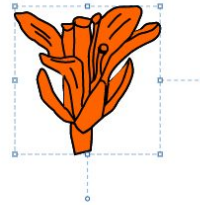
# Early version

- Based off work done by Sakshi Shinghal
- Canvas area for organizing SVG images
- Palette area for selecting them



# Current version

- Palette moved to sidebar
- Side pane displays metadata about document and SVG elements
- Multiple UI functionalities added



Add Text

Clear Canvas

Delete

Toggle Labels

Toggle Colours

Export SVG

Export XML

Group Ungroup

○ □ - 🔍

Elements Document User

Unnamed Element 1

Link

file\_3.bam × ▾

file\_7.bam × ▾

🔗 Edit



# Gene summarization

File	Mean RPKM featureCounts	Mean RPKM HTseq	Median RPKM featureCounts	Median RPKM HTseq	Mean difference	Median difference	Runtime featureCounts	Runtime HTseq	Assigned features featureCounts	Assigned features HTseq	Assigned features difference
3581896	31.44	27.31	1.80	1.58	4.13	0.22	53s	14m26s	20977	20924	53
3581897	17.98	16.87	2.97	2.84	1.11	0.13	55s	14m38s	24202	24129	73
3581898	26.48	23.38	1.83	1.64	3.1	0.19	55s	14m51s	23435	23355	80
3581899	18.30	15.89	3.93	3.49	2.41	0.44	57s	15m03s	24734	24627	107

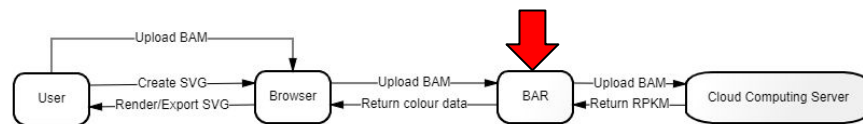
Rewritten to use Rsubread instead of HT-seq

- Faster runtime
- Negligible differences in RPKM values and assigned features

```
1 workflow geneSummarization {
2   File summarizeGenesScript
3   String inputBamFolder
4   File gtf
5
6   call geneSummarizationScript {
7     input:
8     SummarizeGenesScript = summarizeGenesScript,
9     InputBamFolder = inputBamFolder,
10    Gtf = gtf
11  }
12 }
13
14 task geneSummarizationScript {
15   File SummarizeGenesScript
16   String InputBamFolder
17   File Gtf
18
19   command {
20     Rscript ${SummarizeGenesScript} ${InputBamFolder} ${Gtf}
21   }
22   output {
23     File rpkm = "rpkm.txt"
24     File raw_counts_and_lengths = "raw_counts_and_lengths.txt"
25   }
26 }
```

# API connecting front-end and summarization workflow

- Connects cloud server where summarization runs to the front-end
- Receives Google Drive folder ID
- Downloads BAM files from Drive
- Sends to summarization server running on the cloud



# Current functionalities

- Add/remove SVGs from pre-selected list
- Drag, resize and rotate SVG elements
- Upload SVGs from local computer
- Name individual SVG elements, as well as associate samples and citations with them
- Display/hide SVG element labels
- Group/ungroup SVG elements into higher-level elements
- Add different shapes (circles, rectangles, etc.)
- Colour in SVG elements based on the RPKM values from the associated samples
- Display hierarchical structure of SVG elements
- Export SVG file as well as XML file based on the currently added metadata
- Obtain BAM files from a Google Drive folder
- Summarize BAM files to obtain RPKM data



# Future functionalities

- Group/ungroup into different XML <group> tags, rather than just SVG <g> tags
- Fix SVG and XML export so the exported file corresponds to ePlant specifications
- Create gene selection component to specify the gene to be used when displaying fill
- Make API return files rather than file list
- Check summarization status asynchronously
- Check GTF aliases before processing
- Check BAM file integrity/format before processing
- Return single CSV file with all RPKM values for every file, instead of individual files for each
- Rewrite API to use Flask to conform to BAR standards

# Tools used for development

- Node.js
- React.js
- React-bootstrap
- Snap.svg
- Express.js
- Google Drive API
- Rsubread
- HT-seq
- Cromwell