

1 **LARGE-SCALE BIOLOGY ARTICLE**
2

3 **ePlant: Visualizing and Exploring Multiple Levels of Data for
4 Hypothesis Generation in Plant Biology**

5
6 **Jamie Waese¹*, Jim Fan², Asher Pasha¹, Hans Yu¹, Geoff Fucile³, Ruian Shi¹, Matthew
7 Cumming¹, Lawrence Kelley⁴, Michael Sternberg⁴, Vivek Krishnakumar⁵, Erik
8 Ferlanti⁵, Jason Miller⁵, Chris Town⁵, Wolfgang Stuerzlinger⁶ and Nicholas J. Provart^{1*}**

9
10 ¹Department of Cell & Systems Biology/Centre for the Analysis of Genome Evolution and
11 Function, University of Toronto, Toronto, Ontario, Canada

12 ²Department of Computer Science, University of Waterloo, ON. Canada

13 ³SIB Swiss Institute of Bioinformatics, sciCORE Computing Center, University of Basel, Basel,
14 Switzerland.

15 ⁴Imperial College London, London, U.K.

16 ⁵Araport.org / J. Craig Venter Institute, Maryland, U.S.A.

17 ⁶School of Interactive Arts + Technology, Simon Fraser University, BC, Canada

18 *corresponding author: nicholas.provart@utoronto.ca

19
20 **Short title:** ePlant

21
22 **One-sentence summary:** ePlant for hypothesis generation permits the exploration of plant
23 data across >12 orders of magnitude encompassing >20 different kinds of genome-wide
24 data, all in one easy-to-use, open source tool.

25
26 The author responsible for distribution of materials integral to the findings presented in
27 this article in accordance with the policy described in the Instructions for Authors
28 (www.plantcell.org) is: Nicholas J. Provart

29
30 **ABSTRACT**

31 A big challenge in current systems biology research arises when different types of data
32 must be accessed from separate sources and visualized using separate tools. The high
33 cognitive load required to navigate such a workflow is detrimental to hypothesis
34 generation. Accordingly, there is a need for a robust research platform that incorporates all
35 data, and provides integrated search, analysis, and visualization features through a single
36 portal. Here, we present ePlant (<http://bar.utoronto.ca/eplant>), a visual analytic tool for
37 exploring multiple levels of *Arabidopsis* data through a zoomable user interface. ePlant
38 connects to several publicly available web services to download genome, proteome,
39 interactome, transcriptome, and 3D molecular structure data for one or more genes or gene
40 products of interest. Data are displayed with a set of visualization tools that are presented
41 using a conceptual hierarchy from big to small, and many of the tools combine information
42 from more than one data type. We describe the development of ePlant in this paper and
43 present several examples illustrating its integrative features for hypothesis generation. We
44 also describe the process of deploying ePlant as an “app” on Araport. Building on readily
45 available web services, the code for ePlant is freely available for any other biological species
46 research.
47

48 **INTRODUCTION**

49 Many data visualization tools have been created to provide visual depictions of
50 information that is normally not easily visible. Yet, to explore complex phenomena
51 at multiple levels of analysis using existing tools, researchers must visit multiple
52 web sites, each with their own user interface, data input requirements, methods for
53 organizing and categorizing information, and design language for visualizing the
54 particular layer of information that they were created to display.

55 There are many connections between biological entities at different levels of
56 analysis. Figure 1 illustrates the depth and complexity of the relationships between
57 these levels. An investigation of seed development in *Arabidopsis* may focus on one
58 particular transcription factor gene, *ABI3*, but at which level of analysis?

59 Environmental factors can lead to DNA polymorphisms being retained, leading to
60 natural variation in different populations at the level of protein sequences and
61 structures. Subtle variations at these levels can then affect signaling and signal
62 transduction cascades, protein interaction networks, metabolism, subcellular
63 localization, spatio-temporal distribution, and ultimately phenotypic response
64 across various ecotypes.

65 A thorough investigation of *ABI3* would require visiting Araport (Krishnakumar
66 et al., 2014) or TAIR (Berardini et al., 2015) for annotation and sequence
67 information, the BAR *Arabidopsis* Interactions Viewer (Geisler-Lee et al., 2007) or
68 BioGRID (Chatr-Aryamontri et al., 2015) for protein-protein interactions, SUBA3
69 (Tanz et al., 2013) for subcellular localization information, Phyre2 (Kelley et al.,
70 2015) for 3D molecular models, Reactome (Joshi-Tope et al., 2005) for signal,
71 metabolic and gene regulation pathway charts, and dozens of other sites, each
72 dedicated to their own particular layer of analysis.

73 It is difficult to develop hypotheses about complex processes when the
74 information is hard to assemble and laborious to interpret. When researchers must
75 devote a portion of their cognitive load to a computer interface instead of the
76 subjects they are exploring, their “train of thought” will be disrupted and their
77 overall productivity will decrease (Ware, 2012). The effects of interruptions and
78 distractions on cognitive productivity are well described: information overload,

79 increased stress, decreased decision-making accuracy, and the narrowing of
80 attention resulting in the ability to process fewer information cues (Speier et al.,
81 2003). The systems biology research workflow could be improved with the
82 availability of an integrated software platform, with what is known in the
83 information visualization community as a “transparent” user interface, i.e., an
84 interface that “is so easy to use that it all but disappears from consciousness” (Ware,
85 2012).

86 This project addresses these challenges by combining several data visualization
87 tools into the same interface, ordering them into a hierarchy of scale, and providing
88 zoom transitions and integrative connections between the layers so users can
89 explore multiple levels of biological data in new ways (Figure 2). We postulate that
90 applying the principles of user-centered design to build an integrated visual analytic
91 tool for exploring multiple levels of plant data should improve the ability of
92 researchers to extract information from their data, identify connections between
93 layers, facilitate hypothesis generation and, ultimately, promote a deeper
94 understanding of biological processes and functioning.

95 RESULTS

96 We have developed a data integration software tool, ePlant, that not only
97 applies tailored visualizations to more than 10 data types but also integrates data
98 across at least 10 orders of magnitude, from the kilometre scale (natural variation
99 data) to the nanometre scale (protein structure and sequence data), into one, easy-
100 to-use interactive framework. We have developed ePlant based on *Arabidopsis*
101 *thaliana* data, and in this case it taps into > 35 million gene expression
102 measurements, experimentally-documented subcellular localizations for 10,910
103 *Arabidopsis* proteins (with predictions for most of the proteome), ~100,000
104 protein–protein and ~2.7 million protein–DNA interactions, Phyre2-predicted
105 structures covering 23,091 gene products, and 6.19 million non-synonymous SNPs
106 from the 1001 Proteomes website (Supplemental Table 1). In addition, more than a
107 dozen nucleotide-resolution data types (including 100 gigabases of RNA-seq data

108 used to re-annotate the Arabidopsis genome in the Araport 11 release) are also
109 available via Araport's JBrowse instance that has been incorporated into ePlant. We
110 have also created linkages across data scales such that it is possible to ask questions
111 such as, "*Is there a polymorphism that causes a non-synonymous amino acid change*
112 *close to the DNA binding site of my favourite transcription factor?*".

113 **System Architecture and User Interface**

114 ePlant is a collection of programs written with HTML, CSS, JavaScript, and
115 jQuery, bundled together within a custom Zoomable User Interface (ZUI) framework
116 (Figure 2). It is HTML5 compliant and runs within a web browser on most laptops,
117 desktops, and some tablets.

118 ePlant was designed to support data fed dynamically from web services. Upon
119 entering a gene name, alias or AGI ID in the gene selection box in the upper left
120 corner, a data loading management script sends queries to multiple web services
121 (Supplemental Table 1) to retrieve data for each of the ePlant modules. Data are
122 returned asynchronously via AJAX so the program does not freeze while waiting for
123 data to download. Once everything that has been requested has been returned, the
124 data are passed to a function that initializes each module's viewer for each loaded
125 gene.

126 The ePlant user interface has two main elements (Figure 3B): the gene panel
127 and navigation icons on the left; and the module viewer panel on the right. For users
128 who do not know which gene (or genes) they want to look at, the "Expression
129 Angler" button opens a tool that helps identify genes based on a user-defined
130 expression pattern (Austin et al., 2016), and the "Mutant Phenotype Selector" button
131 opens a tool that helps identify genes based on Lloyd and Meinke's mutant
132 phenotype classification system (Lloyd and Meinke, 2012). Both of these features
133 are discussed later in this paper.

134 Downloaded genes appear as rectangular bars in the gene panel. The currently
135 selected gene is coloured green. A vertical stack of icons for selecting the ePlant
136 module to be viewed separates the gene panel from the module viewer panel. The
137 viewers currently include: *Gene Information Viewer, Publication Viewer, Heat Map*

138 *Viewer, World eFP Viewer, Plant eFP Viewer, Tissue & Experiment eFP Viewer, Cell eFP*
139 *Viewer, Chromosome Viewer, Protein Interaction Viewer, Molecule Viewer, Sequence*
140 *Viewer, and Links to External Tools.* Icons appear grey when a module is unavailable,
141 turn black once the data have loaded, and are highlighted green when the module is
142 active (i.e. has been selected for viewing by the user).

143 The module viewer panel shows the content of whichever ePlant module is
144 currently selected. A tab selector at the top of the screen enables users to create
145 multiple views. Beneath the tab selector, a toolbar contains icons for controlling
146 various features, such as *Session history, Screen grab, Zoom in/out, Absolute/relative*
147 *display, Compare genes, Filter data, Custom colour palette, Global/local/custom colour*
148 *gradient, Get citation and experiment information, and Download raw data for the*
149 *currently selected view.* A global options menu in the top right corner of the page
150 allows users to toggle *Zoom transitions, Tooltips, and New user information popups*
151 on and off. There is also an option to *Create a custom URL* that automatically
152 restores the current session upon sharing it with a colleague.

153 Zoomable user interfaces (ZUIs) take advantage of human spatial perception
154 and memory to display more information than could otherwise fit on a desktop
155 computer display. ZUIs have been shown to significantly improve users' ability to
156 find information across multiple layers of data (Helt et al., 2009; Bederson, 2010)
157 and animated navigation transitions have been shown to increase user task
158 performance, even taking into account the time of the transitions themselves (Klein
159 and Bederson, 2005). A custom ZUI framework was created to handle zoom
160 transitions between data visualization modules in ePlant. These transitions do *not*
161 attempt to map spatial and size relationships between the layers. Rather, they
162 produce a "2.5D effect" to indicate a conceptual relationship between the layers.

163 The following ePlant module viewers are organized following a hierarchy of
164 scale from "big" to "small".

165 **Gene Information & Publications Viewer**

166 The Gene Information Viewer (Figure 4A) provides top level access to aliases,
167 full name, description, computational description, curator summary, location, and a

168 visual representation of the gene model structure with intron/exon information
169 about the currently selected gene. It also provides DNA and protein sequences.
170 Immediately beneath it, the Publications Viewer (Figure 4B) provides a list of
171 publications and gene “reference into function” records (GeneRIFs) about the
172 currently selected gene, along with links to the actual papers on PubMed. Both are
173 powered by web services from Araport (Krishnakumar et al., 2014). These modules
174 are atypical for ePlant because they are primarily text based. They were added after
175 user testing feedback.

176 **Heat Map Viewer**

177 The Heat Map Viewer displays all the expression levels for all the samples of all
178 the genes that are loaded, along with the corresponding subcellular localizations of
179 their gene products in one neatly formatted table. This view is near the top of the
180 conceptual hierarchy because it provides a Gestalt sense of the similarities or
181 differences of all the genes/gene products that are loaded. Table cells that are
182 coloured red denote high expression levels (or high levels of confidence in a
183 protein’s subcellular localization), and table cells that are coloured yellow denote
184 low levels. The Heat Map Viewer provides an overview, but determining what
185 sample a red cell corresponds to requires extra mouse-over steps.

186 Figure 5 shows a heat map of twenty-five genes that were identified with the
187 Expression Angler tool for having similar expression patterns as At3g24650 (*ABI3*).
188 This can be quickly confirmed by scanning the heat map to see if the red cells (which
189 represent samples with high expression) are mostly aligned to the same columns, or
190 if there are any obvious outliers. In this case, there are not.

191 When mapping expression levels across several genes, it is important to clarify
192 whether the colour gradient should be determined locally or globally. ePlant can
193 map expression levels locally, which means the colour gradient for each view is
194 determined by the minimum and maximum expression levels of that view. This can
195 be useful to help discern a gene’s expression pattern even if its maximum expression
196 level is significantly lower than other genes being displayed. The “global” colour
197 gradient (default) is useful for comparing gene expression levels, with the minimum

198 and maximum values determined by the lowest and highest expression level of all
199 the genes that have been downloaded. In Figure 5, which uses the “global” colour
200 gradient setting, the maximum expression level for *ABI3* is 1,249 in the stage 9 seeds
201 (as noted in the mouse-over tooltip). Its most similar co-expressed gene is
202 At2g27380 (*EPR1*) with a co-expression coefficient r-value of 0.979 and a maximum
203 expression level of 8,722 (also in the stage 9 seeds). Since *EPR1*’s maximum
204 expression level is nearly seven times higher, *ABI3*’s highest levels are lower on the
205 colour gradient, and thus the heat map cells for the stage 9 seeds appear almost
206 yellow. ePlant also provides a “custom” colour gradient setting so users can map
207 colours to a user-defined threshold.

208 **World eFP Viewer**

209 The World eFP Viewer (Figure 2A) displays natural variation of gene expression
210 levels from 34 ecotypes collected from different parts of the world but grown in a
211 common chamber (Lempe et al., 2005). It draws pin markers (designed to look like a
212 seedling because the data were collected from seedlings) that are coloured
213 according to the expression level for the selected gene for that given ecotype, and
214 placed according to the geographic coordinates of their source on a Google Maps
215 image of the world. Climate data (annual precipitation, maximum temperature, and
216 minimum temperature) is also projected onto the map using the Google Map API
217 raster layer function. Combining ecotype expression data from the Weigel Lab
218 (Lempe et al., 2005) with climate data from the World Bank Climate Portal (Harris et
219 al., 2014) enables researchers to quickly see how ecotypes might differ in response
220 to their environments.

221 This is an update of a similar tool originally included with the Arabidopsis eFP
222 Browser (Winter et al., 2007). The original version used server-side image
223 processing to draw a non-interactive chart, with little concern for data visualization
224 best practices such as “details-on-demand” (Schneiderman, 1996) and “data/ink
225 ratio” (Tufte and Graves-Morris, 1983). A simple task such as determining whether
226 *ABI3* is up- or down-regulated in arid climate regions took considerable effort as the

227 answer required processing information from all over the screen. In the new
228 version, this task can be answered with a single glance.

229 **Plant eFP Viewer**

230 The Plant eFP viewer (Figure 2B, Figure 3B) displays the selected gene's
231 expression pattern by dynamically colouring the tissues of a pictographic
232 representation of a plant according to gene expression levels from multiple
233 experiments. This visualization method is known as an electronic fluorescent
234 pictograph (eFP), and it is a reimplementation of the developmental map view of
235 Arabidopsis eFP Browser by Winter et al. (2007).

236 Several new features have been added. For one, the chart has been redrawn as
237 an SVG image (a vector image instead of a bitmap) and redundant black outlines
238 have been omitted to improve the data/ink ratio (Tufte and Graves-Morris, 1983).
239 Also, since SVG shapes can be filled with any colour programmatically, it is now
240 possible to adjust colour gradients on the fly without having to re-download the
241 image. This makes it possible to toggle between absolute and relative views with
242 almost no latency between screen updates. Finally, expression patterns for several
243 dozen genes occupies much less bandwidth than separate image files for each gene.
244 This makes it possible to switch between eFP images for multiple genes at intervals
245 of 150 ms or faster – a technique known as rapid serial visual presentation (RSVP),
246 discussed later in this paper.

247 Figures 2B and 3B show the spatio-temporal distribution of *ABI3* gene
248 expression levels across the various developmental stages of *Arabidopsis thaliana*
249 based on data from Schmid et al. (2005) and Nakabayashi et al. (2005). With a single
250 glance, it is possible to see that *ABI3* has a narrow expression pattern that is limited
251 to the maturing seeds.

252 **Tissue & Experiment eFP Viewer**

253 The Tissue & Experiment eFP Viewer (Figure 6) provides detail level
254 information about gene expression in individual tissues and the results of
255 perturbation response experiments. The 22 views in this ePlant module display
256 information for 640 separate tissues based on 1385 samples. Multiplying that by the

257 22,814 genes on the ATH1 array (RNA-seq data are available in the Sequence
258 Viewer module) produces a dataset of 31,597,390 records that each query taps into.
259 This is an example of how big data can be explored with a simple graphical interface.

260 Many of the views in this module are based on supplemental views that have
261 been added since the original publication of the Arabidopsis eFP Browser (Winter et
262 al., 2007). They have been updated in keeping with current data visualization best
263 practices. Three new features have been added. First, a vertical stack of thumbnail
264 images along the left side of the window provides a visual method for selecting the
265 active view. Second, the thumbnail images can be sorted either alphabetically or by
266 the maximum expression level for each of the views, making it easy to identify which
267 tissues or experimental conditions are associated with high expression levels for the
268 selected gene. Also, simply by glancing at the proportion of red or yellow in the stack
269 of thumbnail images (a visualization technique sometimes referred to as “small
270 multiples”; Tufte, 1990), it is possible to get a Gestalt sense of the expression pattern
271 for the selected gene in various contexts without opening a single view (e.g., *“Does*
272 *my gene of interest have a narrow or wide expression pattern?”*). Finally, the global
273 colour gradient option (discussed in the Heat Map Viewer section above) is
274 especially useful here because it enables viewers to compare spatio-temporal, tissue
275 specific and perturbation response expression levels all on the same scale (as in
276 Figure 6). The views represent separate experiments, but they are easily
277 comparable because the results are mapped to a common scale. This makes it
278 possible to quickly answer the question, *“In which tissue, and under what*
279 *circumstance, does my gene of interest have the highest expression?”*.

280 As with all the tools in ePlant, the raw data used to generate the charts are
281 available for download as a text file by clicking the “Download Raw Data” button on
282 the toolbar. Table 1 provides a list of all the views available in the Experiment &
283 Tissue eFP Viewer along with their data sources.

284 **Subcellular Localization eFP Viewer**

285 The Subcellular Localization eFP Viewer displays the documented and predicted
286 localization of a gene product within the cell, with a colour gradient representing a

287 confidence score that the selected gene's protein product is found in a given
288 compartment. The data for this module/view originate from SUBA3 database (Tanz
289 et al., 2013) via a web service hosted by the BAR.

290 Like the other eFP viewers in ePlant, the Subcellular eFP Viewer is a
291 reimplementation of an earlier tool by Winter et al. (2007), currently available as a
292 standalone tool at the BAR (http://bar.utoronto.ca/cell_efp/cgi-bin/cell_efp.cgi). As
293 in the original Cell eFP Browser, the numerical score used to compute the shading of
294 each compartment is calculated such that experimentally-determined localizations
295 receive a weighting 5 times that of predicted localizations. Also like the other eFP
296 viewers in ePlant, this updated version uses an SVG image to display the data. One of
297 the advantages of this approach not mentioned previously is that this makes it
298 possible to produce much higher resolution images for publication purposes than
299 the original viewer could generate. Figure 7A shows a screen grab of the Subcellular
300 eFP view for ABI3, while Figure 7B demonstrates how the content can be scaled
301 without image degradation. An example SVG file downloaded from the Plant eFP
302 Viewer for *ABI3* has been included as Supplemental File 1. Such files may be easily
303 used in any vector graphics program to generate high-quality figures. ePlant outputs
304 are freely available under an "open" Creative Commons Attribution license (CC-BY
305 version 4.0).

306 **Chromosome Viewer**

307 The Chromosome Viewer (Figure 2E) provides a pictographic overview of the
308 plant's chromosomes as a series of vertical bars with markers indicating the
309 positions of all the genes that have been downloaded. Spatial relationships within
310 the genome can sometimes indicate functional relationships (Chae et al., 2014;
311 Wisecaver et al., 2017). This viewer can be used to quickly determine the physical
312 location of co-expressed genes, for instance.

313 Clicking on the chromosomes opens a menu listing all the genes at the location
314 that was selected. Since each chromosome contains several thousand genes but the
315 display panel height is typically less than 700 pixels, each pixel represents the
316 location of several genes. This limits the practicality of using this feature as a gene

317 selection method. However, several users reported during user testing that they
318 appreciated how it conveys the sheer number of genes in the genome. Clicking the
319 “thermometer” icon in the toolbar generates a heat map indicating the density of
320 genes within the chromosome. This makes it easy to see if the selected gene is
321 located in a gene-rich region. An annotation tool (accessed by clicking the “pencil”
322 icon) allows users to adjust label colours and sizes in order to make custom charts.

323 **Protein & DNA Interaction Viewer**

324 The Protein & DNA Interaction viewer displays documented and predicted
325 protein-protein interactions (PPIs) and protein-DNA interactions (PDIs) for the
326 selected gene. It uses a node-link charting method, in which the nodes represent
327 proteins or DNA sequences and the links represent interactions between these
328 elements (Figure 8A). This module is a reimplemention of the Arabidopsis
329 Interactions Viewer at the BAR (Geisler-Lee et al., 2007). Several features have been
330 added and the interface has been modified to improve usability.

331 The design of the chart takes advantage of preattentive visual processing
332 (Healey and Enns, 2012) to help users explore multiple levels of data in the same
333 window. DNA elements are drawn as squares and have curved lines to indicate
334 interactions with other proteins. Protein elements are drawn as circles and have
335 straight lines to indicate interactions with other proteins. Interaction line thickness
336 is determined by the interaction confidence value. Line colours are determined by
337 the coexpression coefficient with a yellow-to-red scale. Interactions that have been
338 experimentally determined are drawn with green lines, and clicking them opens a
339 window with the associated paper on PubMed. The borders of protein nodes are
340 coloured according to where each protein is localized within a cell. For instance, a
341 blue border indicates a protein that is mostly found in the nucleus, and an orange
342 border indicates a protein that it is mostly found in the plasma membrane. This
343 makes it possible to quickly answer the question, *“Does my gene of interest mostly*
344 *interact with proteins in the same cell compartment, or across several*
345 *compartments?”*. DNA nodes have black borders because subcellular localization
346 data do not apply to them.

347 The center colour of each node changes from light-grey to dark-grey when the
348 data for that gene have been downloaded. This makes it easy to see if a set of
349 downloaded genes are also interaction partners. Previously, answering this question
350 would require a multi-step list collation process. It can now be done with a glance,
351 simply looking for multiple dark grey node centers.

352 Hovering over the various nodes will open a popup box with annotation
353 information for the gene, and a “Get Data” button that downloads all the data for
354 that gene into ePlant. This makes it possible to “surf” from one gene to another and
355 explore ideas on a whim. Researchers may not initially know which genes to load,
356 but loading one gene could take them on a journey that links to a whole set of genes.

357 In *Arabidopsis*, interacting proteins have an average of 11 interacting partners
358 (Geisler-Lee et al., 2007) but some genes, such as At4g26840, have as many as 172.
359 Drawing that many nodes and links would produce a “hairball” that cannot be easily
360 deciphered. To accommodate these cases, we added a data filtering function that
361 permits users to hide interactions with confidence or correlation values below a
362 customizable threshold. It is also possible to hide either experimentally-determined
363 or predicted PPIs and PDIs.

364 The module was built with JavaScript using Cytoscape.js
365 (<http://js.cytoscape.org>), an open source library for biological network analysis and
366 visualization that is the successor to Cytoscape Web (Lopes et al., 2010). The PPI
367 data come from a database of 70,944 predicted *Arabidopsis* interacting proteins
368 generated by Geisler-Lee et al. (2007) and 36,306 confirmed interaction proteins
369 from the Biomolecular Interaction Network Database (Bader et al., 2003), high-
370 density *Arabidopsis* protein microarrays (Popescu et al., 2007, 2009), Braun et al.'s
371 *Arabidopsis* Interactome (*Arabopsis* Interactome Mapping Consortium, 2011), Wolf
372 Frommer's Membrane protein Interactome Database MIND (Lalonde et al., 2010),
373 and over 1,190 other literature sources. The PDI data come from a database of 1,784
374 confirmed interactions generated by Taylor-Teeple et al. (2015) and DAP-seq data
375 generated by the Ecker lab (O'Malley et al., 2016), which the authors show to be
376 quite similar to ChIP-seq data in terms of quality, while encompassing a far greater
377 number of transcription factors (there are just ~200 *Arabidopsis* ChIP-seq

378 experiments in GEO). All data are downloaded from a web service at the BAR. The
379 interactions in BIND and from other sources were identified using several different
380 methods, such as yeast two hybrid screens, but also via traditional biochemical
381 methods. Subcellular localization data is from SUBA, the Arabidopsis Subcellular
382 Database (Tanz et al., 2013).

383 Figure 8A shows a diagram of protein and DNA interaction partners for
384 At1g54330. This is a good example of how combining multiple levels of data into the
385 same chart can improve systems biology workflows and deepen our understanding
386 of biological functioning, especially in the area of gene regulatory networks.

387 **Molecule Viewer**

388 The Molecule Viewer maps information from four separate databases onto a 3D
389 model of the selected protein's molecular structure. The 3D model (structure
390 models have been computed for 23,091 Arabidopsis gene products as part of the
391 ePlant effort) comes from Phyre2 (Kelley et al., 2015) and data layers include: 1)
392 complete protein sequences from Araport (Krishnakumar et al., 2014); 2) non-
393 synonymous SNP locations in the underlying gene sequence from the 1001
394 Proteomes project (Joshi et al., 2011) with a list of which ecotypes they are found in;
395 3) Pfam domains (Finn et al., 2014); and 4) CDD feature hits (Marchler-Bauer et al.,
396 2015). Drawing this information onto the 3D molecular structure enables
397 researchers to visualize exactly where in the physical model of the protein such
398 features exist. This makes it possible to easily answer the question, *"Is there a*
399 *polymorphism causing a non-synonymous amino acid change near the DNA binding*
400 *site of my favourite transcription factor which might affect its binding to a cis-*
401 *element?"*, as shown in Figure 8B.

402 The PDB file is displayed with JSmol (Hanson et al., 2013). The protein sequence
403 is drawn on the bottom of the screen along with pin markers that indicate the
404 position and frequency of SNP locations. A sliding window enables users to control
405 which part of the sequence they are looking at. Hovering the mouse over the protein
406 sequence highlights the associated location on the 3D model, and hovering the
407 mouse over the 3D model highlights the associated location within the sequence.

408 This enables users to quickly identify which parts of the protein sequence are
409 exposed vs. found in the interior parts of the model, just by moving the mouse over
410 the content. The location of a nsSNP, Pfam domain or CDD feature could have a very
411 large influence on the behaviour of the molecule, and this application of mouseover
412 indicators makes it very easy to find them.

413 **Sequence Viewer**

414 In many ways, sequence browsers were the first zoomable user interfaces for
415 bioinformatics since they enable micro-to-macro level exploration of data, providing
416 detail and overview level information at the same time. They also facilitate
417 comparison of multiple levels of data from a variety of sources (e.g., methylation,
418 phosphorylation, SNPs, conserved non-coding regions, etc.) by mapping each layer
419 onto the chart as a separate track.

420 This module (Figure 2H) uses an implementation of JBrowse (Skinner et al.,
421 2009) using data provided by web services at Araport (Krishnakumar et al., 2014).
422 Due to the complexity of the program, we did not apply the ePlant style guide to this
423 tool so there is some perceptual difference between this and the other ePlant
424 modules. To create more usable screen real estate, the gene panel that occupies the
425 left side of the screen can be slid out of the way by clicking the triangular toggle
426 button at the top of the navigation stack. The Sequence Viewer permits more than a
427 dozen nucleotide-resolution data types (RNA-seq data, conserved non-coding
428 regions, chromatin states, methylation data, non-coding RNAs and others) to be
429 further explored within ePlant.

430 **Links to External Tools**

431 There are many more data visualization modules we would have liked to
432 include with this version of ePlant, but could not for various reasons. The Links to
433 External Tools module contains a list of dynamic links to automatically open the
434 ThaleMine at Araport, TAIR, GeneMANIA, Expressologs, and SeedNet pages for the
435 currently selected gene. While this is not ideal from an integrative tool perspective,
436 it does save many clicks and reduces the inconvenience of having to navigate

437 between sites. This module is easy to update and we plan to add more links in the
438 near future.

439 **Additional ePlant Features**

440 ***Expression Angler***

441 Researchers might not come to ePlant with *a priori* knowledge of which genes
442 they wish to explore. The Expression Angler tool, an implementation of the tool
443 described in Austin et al. (2016), helps users identify and download *Arabidopsis*
444 genes by their expression pattern instead of by name. It does this by calculating the
445 correlation coefficients for expression for all gene expression vectors as compared
446 to an expression pattern that the user defines, or to the expression pattern
447 associated with a single AGI ID or gene name that a user enters (Toufighi et al.,
448 2005). For example, researchers who are interested in exploring the mechanisms
449 associated with seed development may use the Expression Angler to search for
450 genes with high transcript levels in early stage seeds but not in any other tissue.
451 Alternatively, they may use the Expression Angler to find the top 25 genes with
452 similar expression patterns as *ABI3* (as depicted in Figure 5). The tool can be
453 accessed via the button under the gene input box in the upper left corner of the
454 screen.

455 ***Mutant Phenotype Gene Selector***

456 This tool provides two approaches for helping users identify genes associated
457 with loss-of-function mutant phenotypes in *Arabidopsis*: *Search by Classification*, and
458 *Search by Data Table*. It is based on a literature curation effort by Lloyd and Meinke
459 (2012) that includes a database of 2,400 *Arabidopsis* genes with a documented loss-
460 of-function mutant phenotype as well as a proposed schema for categorizing them.
461 The search by classification method uses an interactive Reingold-Tilford tree
462 selection method, implemented with d3.js (<https://d3js.org/>). The search by data
463 table method was built with DataTables (<https://datatables.net/>).

464 ***Rapid Serial Visual Presentation***

465 Identifying genes of interest from a large set of eFP images can be a daunting
466 task. To succeed, researchers must find subtle differences between multiple nearly
467 identical images. The Rapid Serial Visual Presentation (RSVP) display technique has
468 the potential to improve the experience as it exploits our ability to recognize
469 differences between images when they are displayed on a screen in a rapid and
470 serial manner. The technique is known to be an efficient way to find the presence or
471 absence of a specific item within a set of images (Spence, 2002), akin to flipping
472 through a book to find a specific picture.

473 The “Slide Show” RSVP feature, accessed near the top of the gene panel,
474 automatically advances the currently selected gene every 250 milliseconds. Upon
475 reaching the bottom of the list it cycles back to the top. The “Hover” RSVP feature
476 enables users to hover their mouse over the gene panel to adjust the currently
477 selected gene. Moving the mouse up and down over the list produces a “user
478 controlled” RSVP effect. We have shown through controlled user-testing that both of
479 these methods are more efficient than “Point & Click” when it comes to selecting
480 genes of interest from a set of eFP images (Waese et al., 2016).

481 **DISCUSSION**

482 ePlant permits researchers to easily see where and when a gene is “active”, how
483 its protein product can fold into a molecular machine to do what it needs to do and
484 whether there are any natural genetic variants of the gene that might allow it to do it
485 better. It is an open source visual analytic platform that was designed to help plant
486 researchers seamlessly explore data from different biological levels through a single
487 window. It uses a zoomable user interface that enables users to quickly transition
488 from natural variation at the kilometer scale, through gene expression levels in
489 tissues and cell types, subcellular localization of gene products, protein–protein and
490 protein–DNA interactors, to protein tertiary structure and gene sequences at the
491 nanometer scale.

492 Integrating data from different biological levels can allow novel hypotheses to
493 be generated. By combining data from several biological levels of analysis into the
494 same view, ePlant makes it possible to easily examine protein–protein interactions
495 and ask whether these protein products are in the same compartment, what the
496 tertiary structure of a protein product might be and whether there are any
497 polymorphisms that lie close to structurally-important features, like DNA-binding
498 sites. Adams et al. (2017) have shown that structural clustering of variation can
499 predict functional sites in proteins. Our lab is interested in natural variation in ABA
500 signaling and an analysis of ABA-related bZIP transcription factors (*ABSCISIC ACID-*
501 *RESPONSIVE ELEMENT BINDING FACTORS 1* through *4* and *ABI5*, collapsed to a
502 consensus sequence) show that there are a few frequent variants close to the DNA
503 binding site (**Supplemental Figure 1**). The representation of variant frequency
504 across ecotypes by pin size in the ePlant Molecule Viewer is also helping us
505 prioritize which variants to focus our analysis on – there are many non-synonymous
506 SNPs that occur in just one ecotype that we hypothesize to be of less functional
507 importance than those that are found in several ecotypes – this sort of analysis was
508 not possible with the prototypic ePlant interface released several years ago (Fucile
509 et al., 2011). Not having to switch windows and contend with several different user
510 interfaces to explore an idea from multiple perspectives should liberate researchers
511 and make it easier to stay “on task” and make creative associations without
512 distraction.

513 ePlant was developed with regard to best practices for user experience design
514 and data visualization, as well as with feedback gathered from two rounds of user
515 testing (see **Methods**). We have successfully deployed ePlant on the new
516 international portal for Arabidopsis information, Araport.org (Krishnakumar et al.,
517 2014). This is a large collaborative effort that demonstrates the power of a federated
518 web service-based approach in integrating and visualizing data from multiple
519 sources, as articulated by the International Arabidopsis Informatics Consortium
520 (2012). We have made the project open source, such that other groups may develop
521 modules for ePlant as new data types become available and new linkages between
522 different levels of data are discovered.

523 We have received funding from Genome Canada to leverage the ePlant
524 framework to create 15 ePlants for agronomically-important species including
525 tomato, maize, wheat, and soybean. Here a novel “navigator” will be developed to
526 readily permit the exploration of homologous sequences and their associated
527 transcriptomic, proteomics, structural and other data. This framework would be
528 highly useful to improve crop species, and being able to efficiently query and
529 visualize the huge amount of data generated in the past five years will be key to
530 improving and managing these crops to feed, shelter and power a world of 9 billion
531 people by the year 2050. By adding multiple species to the framework (through a
532 pipeline that is also being developed as part of this grant), it will be possible to see if
533 non-synonymous changes map to the same location in one protein’s structure as do
534 non-synonymous changes in another species for a homologous gene. If that is the
535 case, then the likelihood of that polymorphism being biologically relevant would be
536 high. Other powerful research-driven questions that would be possible to ask with
537 this interface include, *“Which homolog of my gene of interest in the species I work on*
538 *has the same expression pattern in equivalent tissues as in the reference species?”*.
539 These kinds of questions are relevant for translational biology, that is, for extending
540 the information and knowledge derived from a reference species into
541 agronomically-important ones. In principle, the benefits of our systems approach
542 extend to any species with available genomic sequences.

543 METHODS

544 System Architecture

545 The system architecture (Figure 3A) can be divided into five categories: 1)
546 databases; 2) web services; 3) data processing functions; 4) ZUI framework; and 5)
547 data visualization modules. ePlant aggregates data from numerous sources, most of
548 which are stored in SQL databases on servers hosted by Araport, TAIR, or our own
549 BAR (Toufighi et al., 2005). The actual data are accessed via web services (typically
550 served up by Perl or Python CGI scripts) hosted on the same server as the databases.
551 When a user selects a gene to download, ePlant sends a batch of queries to each of

552 the web services associated with the various data visualization modules. They
553 return several file formats depending on the nature of the data: JSON objects, XML
554 files, and pre-rendered PNG images (in the case of the Gene Cloud images, Krouk et
555 al., 2015). Databases and web services (i.e., categories one and two in Figure 3A) are
556 server side constructs and can be considered “back-end” components of ePlant.

557 Often, data must be reformatted and processed before they can be visualized.
558 Hard coding the myriad number of data permutations (i.e., category three in Figure
559 3A) directly into the visualization modules would be difficult to maintain if the data
560 format changed or if new data sources become available. Thus, although data
561 processing functions are executed locally on the client machine, they are separate
562 elements within the system architecture. These functions can be considered a
563 “middle layer”.

564 On the “front-end”, ePlant maintains separate functions for ZUI management
565 and data visualization (i.e., categories four and five in Figure 3A). To maximize
566 interface responsiveness, these functions are executed locally on the client machine.
567 The ZUI framework is responsible for drawing interface elements to the screen and
568 triggering various functions in response to user input. The data visualization
569 modules consist of separate programs that are initialized when data becomes
570 available and then run in the background, waiting for the ZUI framework to make
571 their screen visible.

572 For the data visualization modules that are based on HTML5 canvas, the ZUI
573 framework treats each module as a separate <div> and simply animates the scale
574 and visibility of that element. For the other modules, the zoom transitions are built
575 into the views themselves. For example, the eFP viewers use CSS transitions defined
576 within their own function scopes, and the Molecule Viewer calls the JSmol library to
577 resize the 3D molecule model.

578 ePlant was written to be easily expandable. Adding new data visualization
579 modules is a simple matter of adding the necessary data loading and visualization
580 programs to the host directory, adding citation and data source information, and
581 adding an icon to the ZUI navigation panel. The code is well documented and

582 available as open source code on GitHub for anyone to explore or fork and modify
583 (see below).

584 **User Testing**

585 To ensure the relevance and usefulness of ePlant for its intended users, we
586 adopted an “agile” approach to software design (Highsmith and Cockburn, 2001), a
587 process that includes frequent user testing, analysis of user needs, prototyping and
588 refinement. As part of that process, we conducted two rounds of user testing at the
589 2014 and 2015 International Conference of Arabidopsis Research (ICAR). Attendees
590 were invited to follow a user testing protocol based on Nielsen’s guidelines for
591 usability engineering (1993) and (Hudson, 2014) that consisted of three phases:
592 free exploration of the tool; completion of ten sample tasks; and a Google Forms
593 questionnaire. The protocol was approved by the University of Toronto Research
594 Ethics Board (Protocol #30490).

595 Participants were recorded with Screencast-O-Matic software as they interacted
596 with the tool. All mouse clicks and verbal comments were recorded, and participants
597 were asked to “think out loud” so we could collect qualitative feedback about
598 interactions as they happened. Performance time was measured for typical tasks
599 such as: *Which tissue is ABI3 most strongly expressed in? Where is ABI3 localized in*
600 *the cell? Can you name an interaction partner for ABI3? In what part of the world does*
601 *AT1G16850 show the most natural variation of expression? What is the annotation for*
602 *ATAP3?* The same tasks were presented in the same order across both user testing
603 sessions. Tasks that could not be answered by the majority of participants in twenty
604 seconds or less were flagged for additional development.

605 At the 2014 ICAR in Vancouver, thirteen participants completed the study (4
606 professors, 2 post docs, 5 PhD candidates, and 2 industry researchers). At the 2015
607 ICAR in Paris, eighteen participants completed the study (8 professors, 1 post doc, 3
608 PhD candidates, 3 industry researchers, and 3 undeclared). This may not seem like a
609 large number, however Nielsen and Landauer (1993) found that the typical user
610 testing session will identify 31% of all the usability problems in a design, and that
611 85% of a site’s problems can be found with as few as five participants.

612 After using the tool for about ten minutes, participants were asked to complete
613 a Google Forms questionnaire with a 7-point Likert scale with the following
614 questions: 1) *Please rate the quality of ePlant's user interface*; 2) *Please rate how*
615 *useful ePlant is for Arabidopsis researchers*; 3) *How would you describe the depth of*
616 *information contained in ePlant?*; 4) *How would you compare ePlant against current*
617 *methods for accessing the same information?*; 5) *How likely are you to use ePlant in a*
618 *research project?*; 6) *How would you describe the depth of information contained in*
619 *ePlant?*; 7) *How likely are you to use ePlant again?*; and 8) *Please rate your overall*
620 *user experience of using ePlant.*

621 Responses across both years were positive. In 2015 almost all participants
622 responded with the most positive response for "*How likely are you to use ePlant in a*
623 *research project?*" and "*How likely are you to use ePlant again?*" These are essentially
624 the same question, and the response suggests that ePlant successfully delivers on
625 the objective to build a research platform that plant biologists want to use.

626 Quantitative data provides a snapshot of the efficacy of the tool; however, the
627 main value from user testing is found in the qualitative data that was collected.
628 Notes taken while coding the screencasts produced a total of 88 feature requests,
629 bug reports, interface modifications, and other suggestions on how to improve the
630 final tool. These notes were entered into an issue-tracking platform called Pivotal
631 Tracker that allows tasks to be sorted according to difficulty, and assigned to
632 individual programmers to work on. At this time, virtually all of the tasks have been
633 addressed and/or implemented.

634 **Implementation on Araport**

635 ePlant was initially written as a standalone program for the Bio-Analytic
636 Resource for Plant Biology (<http://bar.utoronto.ca/>). It has been deployed as a
637 science app on Araport, accessible from Araport's front page at Araport.org. Using
638 Araport's Yeoman-based application scaffold (called aip-science-app), ePlant front-
639 end code was ported and integrated into the Araport science app framework. Two
640 multi-point pass-through Araport Data And Microservices API (ADAMA) adapters,
641 "eplant_service" and "expression_angler_service", were developed to retrieve data

642 from the BAR web services. These adapters, hosted on a public GitHub repository
643 (https://github.com/BioAnalyticResource/Araport_ePlant_ADAMA), were
644 registered with Araport as community API adapters. Araport users may try out
645 these adapters at <https://www.araport.org/api-explorer> after signing into Araport
646 or via the command line using an Araport OAuth 2.0 access token
647 (<https://www.araport.org/docs/building-community-apis-adama/getting-token>).
648 The JavaScript code of ePlant was modified to load data using these ADAMA
649 adapters. The modified code is hosted at this public GitHub repository:
650 https://github.com/BioAnalyticResource/Araport_ePlant. Users may run ePlant on
651 their computers using Araport's text environment built with Grunt and Node.js.
652 They may also deploy the ePlant app into their own Araport workspaces.

653 **SUPPLEMENTAL DATA**

654 **Supplemental Figure 1.** Analysis of nsSNPs near to ABF1's DNA binding site
655 performed using ePlant, or across ABA-related bZIP transcription factors using data
656 from the 1001 Proteomes site (Supports Figure 8B).

657 **Supplemental Table 1.** A list of web services used by ePlant to populate the data
658 various data visualization modules.

659 **Supplemental File 1.** An example SVG file downloaded from the Plant eFP Viewer
660 for *ABI3*.

661 **ACKNOWLEDGEMENTS**

662 This Project was funded by the Government of Canada through Genome Canada and
663 Ontario Genomics (OGI-071). We thank the reviewers for helpful suggestions.

664 **AUTHOR CONTRIBUTIONS**

665 J.W. conceived, programmed, and coordinated development of ePlant. J.F.
666 programmed ePlant. A.P. developed web services for ePlant, and deployed ePlant on
667 Araport. H.Y. programmed initial ePlant framework. G.F. developed components of
668 ePlant's Molecule Viewer. R.S. developed ePlant's Interaction Viewer. M.C.
669 performed nsSNP analysis. L.K. and M.S. provided Phyre2 structure-ome. V.K., E.F.,
670 J.M. and C.T. developed Araport components for the ePlant app. W.S. provided input
671 on ePlant user testing and user interface. N.J.P. conceived and oversaw overall
672 ePlant project. J.W. and N.J.P. wrote most of the manuscript, with a small amount of
673 additional material provided by some of the coauthors.

674

675 **REFERENCES**

- 676 Arabidopsis Interactome Mapping Consortium (2011). Evidence for Network Evolution
677 in an Arabidopsis Interactome Map. *Science* 333, 601–607.
- 678 Adams, J., Mansfield, M.J., Richard, D.J., and Doxey, A.C. (2017). Lineage-specific
679 mutational clustering in protein structures predicts evolutionary shifts in function.
680 *Bioinformatics*. doi: 10.1093/bioinformatics/btw815.
- 681 Arabidopsis Interactome Mapping Consortium (2011). Evidence for Network
682 Evolution in an Arabidopsis Interactome Map. *Science* 333, 601–607. Austin, R.S.,
683 Hiu, S., Waese, J., Ierullo, M., Pasha, A., Wang, T., Fan, J., Foong, C., Breit, R., Desveaux,
684 D., et al. (2016). New BAR Tools for Mining Expression Data and Exploring Cis-
685 Elements in *Arabidopsis thaliana*. *Plant J.* 88, 490–504.
- 686 Bader, G.D., Betel, D., and Hogue, C.W.V. (2003). BIND: the Biomolecular Interaction
687 Network Database. *Nucleic Acids Res.* 31, 248–250.
- 688 Bederson, B.B. (2010). The Promise of Zoomable User Interfaces. In *Proceedings of*
689 *the 3rd International Symposium on Visual Information Communication*, (New York,
690 NY, USA: ACM). doi: 10.1145/1865841.1865844.

- 691 Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E.
692 (2015). The Arabidopsis Information Resource: Making and mining the “gold
693 standard” annotated reference plant genome. *Genesis* 53, 474–485.
- 694 Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and
695 Benfey, P.N. (2003). A Gene Expression Map of the Arabidopsis Root. *Science* 302,
696 1956–1960.
- 697 Böhmer, M., and Schroeder, J.I. (2011). Quantitative transcriptomic analysis of
698 abscisic acid-induced and reactive oxygen species-dependent expression changes
699 and proteomic profiling in *Arabidopsis* suspension cells. *Plant J.* 67, 105–118.
- 700 Casson, S., Spencer, M., Walker, K., and Lindsey, K. (2005). Laser capture
701 microdissection for the analysis of gene expression during embryogenesis of
702 *Arabidopsis*. *Plant J.* 42, 111–123.
- 703 Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S.Y. (2014). Genomic signatures of
704 specialized metabolism in plants. *Science* 344, 510–513.
- 705 Chatr-Aryamontri, A., Breitkreutz, B.-J., Oughtred, R., Boucher, L., Heinicke, S., Chen,
706 D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., et al. (2015). The BioGRID
707 interaction database: 2015 update. *Nucleic Acids Res.* 43, D470–478.
- 708 Couldridge, C., Newbury, H.J., Ford-Lloyd, B., Bale, J., and Pritchard, J. (2007).
709 Exploring plant responses to aphid feeding using a full *Arabidopsis* microarray
710 reveals a small number of genes with significantly altered expression. *Bull. Entomol.*
711 *Res.* 97, 523–532.
- 712 Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A.,
713 Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families
714 database. *Nucleic Acids Res.* 42, D222–D230.
- 715 Fucile, G., Di Biase, D., Nahal, H., La, G., Khodabandeh, S., Chen, Y., Easley, K.,
716 Christendat, D., Kelley, L., Provart, N.J. (2011) ePlant and the 3D Data Display
717 Initiative: Integrative Systems Biology on the World Wide Web. *PLoS ONE* 6:
718 e15237.
- 719 Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M.
720 (2007). A predicted interactome for *Arabidopsis*. *Plant Physiol.* 145, 317–329.
- 721 Gilding, E.K., and Marks, M.D. (2010). Analysis of purified glabra3-shapeshifter
722 trichomes reveals a role for NOECK in regulating early trichome morphogenic
723 events. *Plant J.* 64, 304–317.
- 724 Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W.,
725 Ogawa, M., Yamauchi, Y., Preston, J., Aoki, K., et al. (2008). The AtGenExpress

- 726 hormone and chemical treatment data set: experimental design, data evaluation,
727 model data analysis and data access. *Plant J.* 55, 526–542.
- 728 Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., and Sussman, J.L. (2013). JSmol and
729 the Next-Generation Web-Based Representation of 3D Molecular Structure as
730 Applied to Proteopedia. *Isr. J. Chem.* 53, 207–216.
- 731 Harris, I., Jones, P. d., Osborn, T. j., and Lister, D. h. (2014). Updated high-resolution
732 grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* 34,
733 623–642.
- 734 Healey, C.G., and Enns, J.T. (2012). Attention and visual memory in visualization and
735 computer graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 1170–1188.
- 736 Helt, G.A., Nicol, J.W., Erwin, E., Blossom, E., Blanchard, S.G., Chervitz, S.A., Harmon,
737 C., and Loraine, A.E. (2009). Genoviz Software Development Kit: Java tool kit for
738 building genomics visualization applications. *BMC Bioinformatics* 10, 266.
- 739 Highsmith, J., and Cockburn, A. (2001). Agile software development: the business of
740 innovation. *Computer* 34, 120–127.
- 741 Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male
742 gametophyte development in *Arabidopsis*. *Genome Biol.* 5, R85.
- 743 Hudson, W. (2014). Agile UX and UCD. In CHI '14 Extended Abstracts on Human
744 Factors in Computing Systems, (New York, NY, USA: ACM), pp. 1009–1010.
- 745 International Arabidopsis Informatics Consortium (2012). Taking the Next Step:
746 Building an *Arabidopsis* Information Portal. *Plant Cell* 24, 2248–2256.
- 747 Joshi, H.J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt,
748 Schulze, W.X., Sun, Q., Wijk, K.J. van, Egelhofer, V., et al. (2011). MASCP Gator: An
749 Aggregation Portal for the Visualization of *Arabidopsis* Proteomics Data. *Plant
750 Physiol.* 155, 259–270.
- 751 Joshi-Tope, G., Gillespie, M., Vastraik, I., D'Eustachio, P., Schmidt, E., Bono, B. de, Jassal,
752 B., Gopinath, G.R., Wu, G.R., Matthews, L., et al. (2005). Reactome: a knowledgebase
753 of biological pathways. *Nucleic Acids Res.* 33, D428–D432.
- 754 Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The
755 Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10,
756 845–858.
- 757 Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C.,
758 Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress
759 expression data set: protocols, evaluation and model data analysis of UV-B light,
760 drought and cold stress responses. *Plant J.* 50, 347–363.

- 761 Klein, C., and Bederson, B.B. (2005). Benefits of Animated Scrolling. In CHI '05
762 Extended Abstracts on Human Factors in Computing Systems, (New York, NY, USA:
763 ACM), pp. 1965–1968.
- 764 Krishnakumar, V., Hanlon, M.R., Contrino, S., Ferlanti, E.S., Karamycheva, S., Kim, M.,
765 Rosen, B.D., Cheng, C.-Y., Moreira, W., Mock, S.A., et al. (2014). Araport: the
766 Arabidopsis Information Portal. *Nucleic Acids Res.* 43, D1003-D1009.
- 767 Lalonde, S., Sero, A., Pratelli, R., Pilot, G., Chen, J., Sardi, M.I., Parsa, S.A., Kim, D.-Y.,
768 Acharya, B.R., Stein, E.V., et al. (2010). A membrane protein/signaling protein
769 interaction network for Arabidopsis version AMPv2. *Front. Physiol.* doi.
770 10.3389/fphys.2010.00024.
- 771 Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., and Weigel, D.
772 (2005). Diversity of Flowering Responses in Wild Arabidopsis thaliana Strains. *PLoS*
773 *Genet.* 1, e6.
- 774 Lloyd, J., and Meinke, D. (2012). A Comprehensive Dataset of Genes with a Loss-of-
775 Function Mutant Phenotype in Arabidopsis. *Plant Physiol.* 158, 1115–1129.
- 776 Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q., and Bader, G.D. (2010).
777 Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26, 2347–
778 2348.
- 779 Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y.,
780 Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved
781 domain database. *Nucleic Acids Res.* 43, D222-226.
- 782 Marks, M.D., Wenger, J.P., Gilding, E., Jilk, R., and Dixon, R.A. (2009). Transcriptome
783 Analysis of Arabidopsis Wild-Type and gl3-sst sim Trichomes Identifies Four
784 Additional Genes Required for Trichome Development. *Mol. Plant* 2, 803–822.
- 785 Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005).
786 Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination:
787 epigenetic and genetic regulation of transcription in seed. *Plant J.* 41, 697–709.
- 788 Nawy, T., Lee, J.-Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K.,
789 and Benfey, P.N. (2005). Transcriptional Profile of the Arabidopsis Root Quiescent
790 Center. *Plant Cell* 17, 1908–1925.
- 791 Nielsen, J. (1993). Usability Engineering (San Francisco, CA, USA: Morgan Kaufmann
792 Publishers Inc.).
- 793 Nielson, J., and Landauer, T. A mathematical model of the finding of usability
794 problems. *Proc. ACM INTERCHI93 Conf. Amst. Neth. 24-29 April 1993*, 206–213.

- 795 O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M.,
796 Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the
797 Regulatory DNA Landscape. *Cell* 165, 1280–1292.
- 798 Pandey, S., Wang, R.-S., Wilson, L., Li, S., Zhao, Z., Gookin, T.E., Assmann, S.M., and
799 Albert, R. (2010). Boolean modeling of transcriptome data reveals novel modes of
800 heterotrimeric G-protein action. *Mol. Syst. Biol.* 6, 372.
- 801 Pillitteri, L.J., Peterson, K.M., Horst, R.J., and Torii, K.U. (2011). Molecular Profiling of
802 Stomatal Meristemoids Reveals New Component of Asymmetric Cell Division and
803 Commonalities among Stem Cell Populations in *Arabidopsis*. *Plant Cell* 23, 3260–
804 3275.
- 805 Popescu, S.C., Snyder, M., and Dinesh-Kumar, S.P. (2007). *Arabidopsis* Protein
806 Microarrays for the High-Throughput Identification of Protein-Protein Interactions.
807 *Plant Signal. Behav.* 2, 416–420.
- 808 Popescu, S.C., Popescu, G.V., Bachan, S., Zhang, Z., Gerstein, M., Snyder, M., and
809 Dinesh-Kumar, S.P. (2009). MAPK target networks in *Arabidopsis thaliana* revealed
810 using functional protein microarrays. *Genes Dev.* 23, 80–92.
- 811 Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson,
812 M.A., and Palanivelu, R. (2009). Penetration of the stigma and style elicits a novel
813 transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS
814 Genet.* 5, e1000621.
- 815 Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B.,
816 Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana*
817 development. *Nat. Genet.* 37, 501–506.
- 818 Schneiderman, B. (1996). The eyes have it: A task by data type taxonomy for
819 information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual
820 Languages*, (IEEE), pp. 336–343.
- 821 Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., and Holmes, I.H. (2009). JBrowse:
822 A next-generation genome browser. *Genome Res.* 19, 1630–1638.
- 823 Speier, C., Vessey, I., and Valacich, J.S. (2003). The Effects of Interruptions, Task
824 Complexity, and Information Presentation on Computer-Supported Decision-Making
825 Performance. *Decis. Sci.* 34, 771–797.
- 826 Spence, R. (2002). Rapid, Serial and Visual: A Presentation Technique with Potential.
827 *Inf. Vis.* 1, 13–19.
- 828 Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F.
829 (2005). Cuticular Lipid Composition, Surface Structure, and Gene Expression in
830 *Arabidopsis* Stem Epidermis. *Plant Physiol.* 139, 1649–1665.

- 831 Swanson, R., Clark, T., and Preuss, D. (2005). Expression profiling of Arabidopsis
832 stigma tissue identifies stigma-specific genes. *Sex. Plant Reprod.* 18, 163–171.
- 833 Tanz, S.K., Castleden, I., Hooper, C.M., Vacher, M., Small, I., and Millar, H.A. (2013).
834 SUBA3: a database for integrating experimentation and prediction to define the
835 SUBcellular location of proteins in Arabidopsis. *Nucleic Acids Res.* 41, D1185-1191.
- 836 Taylor-Teeple, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young,
837 N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., et al. (2015). An Arabidopsis gene
838 regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575.
- 839 Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J. (2005). The Botany Array
840 Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J.* 43, 153–
841 163.
- 842 Tufte, E.R. (1990). *Envisioning Information* (Graphics Press, Cheshire, CT).
- 843 Tufte, E.R., and Graves-Morris, P.R. (1983). *The visual display of quantitative
844 information* (Graphics Press, Cheshire, CT).
- 845 Waese, J., Stuerzlinger, W., and Provart, N.J. (2016). An Evaluation of Interaction
846 Methods for Controlling RSVP Displays in Visual Search Tasks. In *Big Data Visual
847 Analytics (BDVA)*, 2016, (Sydney: IEEE Xplore). doi: 10.1109/BDVA.2016.7787041.
- 848 Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D., Fu, X.-D., and
849 Dong, X. (2011). Timing of plant immune responses by a central circadian regulator.
850 *Nature* 470, 110–114.
- 851 Ware, C. (2012). *Information Visualization: Perception for Design* (Elsevier).
- 852 Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An
853 “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-
854 Scale Biological Data Sets. *PLoS ONE* 2, e718.
- 855 Wisecaver, J.H., Borowsky, A.T., Tzin, V., Jander, G., Kliebenstein, D.J., and Rokas, A.
856 (2017). A Global Co-expression Network Approach for Connecting Genes to
857 Specialized Metabolic Pathways in Plants. *Plant Cell* tpc.00009.2017.
- 858 Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V. (2009). Gene expression map
859 of the Arabidopsis shoot apical meristem stem cell niche. *Proc. Natl. Acad. Sci. U. S. A.*
860 106, 4941–4946.
- 861 Yang, Y., Costa, A., Leonhardt, N., Siegel, R.S., and Schroeder, J.I. (2008). Isolation of a
862 strong Arabidopsis guard cell promoter and its potential as a research tool. *Plant
863 Methods* 4, 6.
- 864

865 **FIGURE LEGENDS**

866 **Figure 1.** Connections between biological entities at different levels of analysis.

867

868 **Figure 2.** ePlant's module viewers (each displaying data for *ABI3*) are presented
869 here with the intention of illustrating ePlant's hierarchy of scale. To see detail, view
870 online version. **(A)** World eFP Viewer, **(B)** Plant eFP Viewer, **(C)** Tissue &
871 Experiment eFP Viewer, **(D)** Subcellular eFP Viewer, **(E)** Chromosome Viewer, **(F)**
872 Interaction Viewer, **(G)** Molecule Viewer, **(H)** Sequence Viewer. High resolution
873 vector graphics are available for most views (see example in Supplemental File 1).
874 ePlant outputs are freely usable via an "open" license.

875

876 **Figure 3.** ePlant system design and user interface. **(A)** ePlant system architecture.
877 **(B)** ePlant user interface showing the expression pattern of *ABI3* with the Plant eFP
878 Viewer. See detail in online version of figure.

879

880 **Figure 4.** ePlant views providing information from Araport. **(A)** ePlant Gene
881 Information Viewer. **(B)** ePlant Publications Viewer.

882

883 **Figure 5.** The Heat Map Viewer showing 350+ expression level samples for twenty-
884 five genes identified with the Expression Angler for having similar expression
885 patterns to *ABI3* (At3g24650). The "global" colour gradient is selected, making it
886 easy to see the variability in the expression levels of the various genes.

887

888 **Figure 6.** Six of more than twenty views from the Tissue & Experiment eFP Viewer.
889 Each view displays expression levels for *ABI3* with the "custom" colour gradient
890 setting, with red = 100 expression units: **(A)** Root, **(B)** Guard and Mesophyll Cells,
891 **(C)** Microgametogenesis, **(D)** Biotic Stress: *Pseudomonas syringae*, **(E)** Abiotic Stress,
892 **(F)** Pollen Germination. Some views are truncated for display here; see online
893 version of figure to be able to see detail. ePlant outputs for all views may be

894 downloaded as high-resolution vector graphic files and are freely available for use
895 by any researcher under an “open” license.

896

897 **Figure 7.** The Subcellular Localization eFP Viewer. **(A)** ABI3 is mostly localized in
898 the nucleus. **(B)** An inset of a high-resolution version of the same image.

899

900 **Figure 8.** ePlant Interaction and Molecule Viewers. **(A)** Protein and DNA Interaction
901 Viewer showing interactions for At1g54330. **(B)** Molecule Viewer showing the
902 transcription factor ABI3’s Phyre2-predicted partial 3D structure with its DNA
903 binding site highlighted in blue, and two non-synonymous changes (from a web
904 service provided by the 1001 Proteomes site) highlighted in green. Changes of
905 higher frequency are denoted by larger, redder pins above the sequence below.

906

907 **Table 1.** A list of Experimental & Tissue eFP views and data sources.

VIEW	DATA SOURCE
1. Abiotic Stress	(Kilian et al., 2007)
2. Biotic Stress – <i>Botrytis cinerea</i>	AtGenExpress initiative
3. Biotic Stress – Elicitors	AtGenExpress initiative
4. Biotic Stress – <i>Erysiphe orontii</i>	AtGenExpress initiative
5. Biotic Stress – <i>Hyaloperonospora arabidopsis</i>	(Wang et al., 2011)
6. Biotic Stress – <i>Myzus persicae</i>	(Couldridge et al., 2007)
7. Biotic Stress – <i>Phytophthora infestans</i>	AtGenExpress initiative
8. Biotic Stress – <i>Pseudomonas syringae</i>	AtGenExpress initiative
9. Chemical	(Goda et al., 2008)
10. Guard Cell – Meristemoids	(Pillitteri et al., 2011)
11. Guard Cell – Mutant and Wild Type Guard Cell ABA Response	(Pandey et al., 2010)
12. Guard Cell – Suspension Cell ABA Response with ROS Scavenger	(Böhmer and Schroeder, 2011)
13. Tissue Specific – Embryo Development	(Casson et al., 2005)
14. Tissue Specific – Guard and Mesophyll Cells	(Yang et al., 2008)
15. Tissue Specific – Microgametogenesis	(Honys and Twell, 2004)
16. Tissue Specific – Pollen Germination	(Qin et al., 2009)
17. Tissue Specific – Root	(Birnbaum et al., 2003) (Nawy et al., 2005)
18. Tissue Specific – Shoot Apical Meristem	(Yadav et al., 2009)

-
- 19.** Tissue Specific – Stem Epidermis (Suh et al., 2005)
-
- 20.** Tissue Specific - Stigma and Ovaries (Swanson et al., 2005)
-
- 21.** Tissue Specific – Trichomes (Gilding and Marks, 2010; Marks et al., 2009)
-
- 22.** Tissue Specific – Xylem and Cork (NASCArrays experiment #92)
-

908

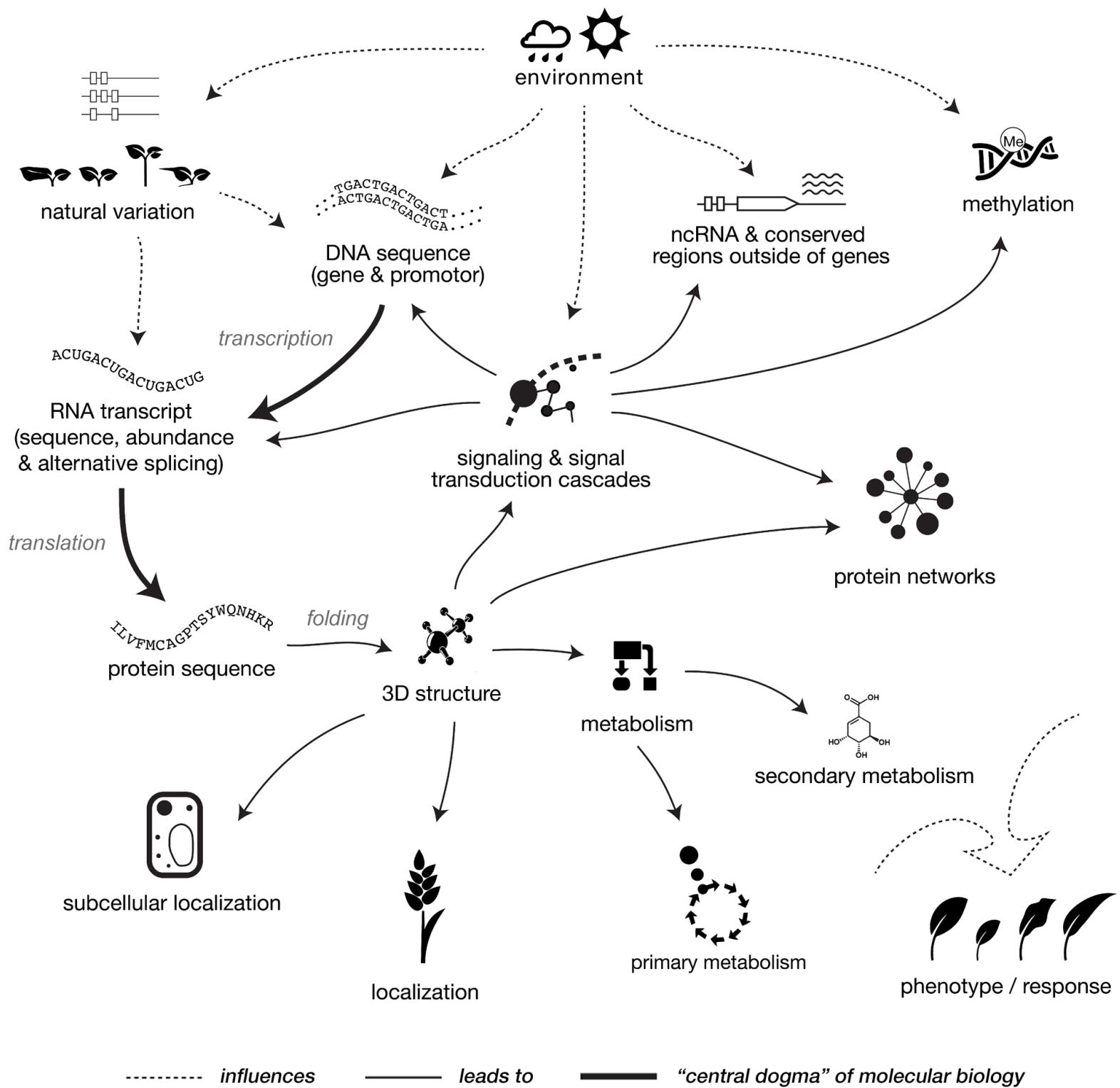


Figure 1. Connections between biological entities at different levels of analysis.

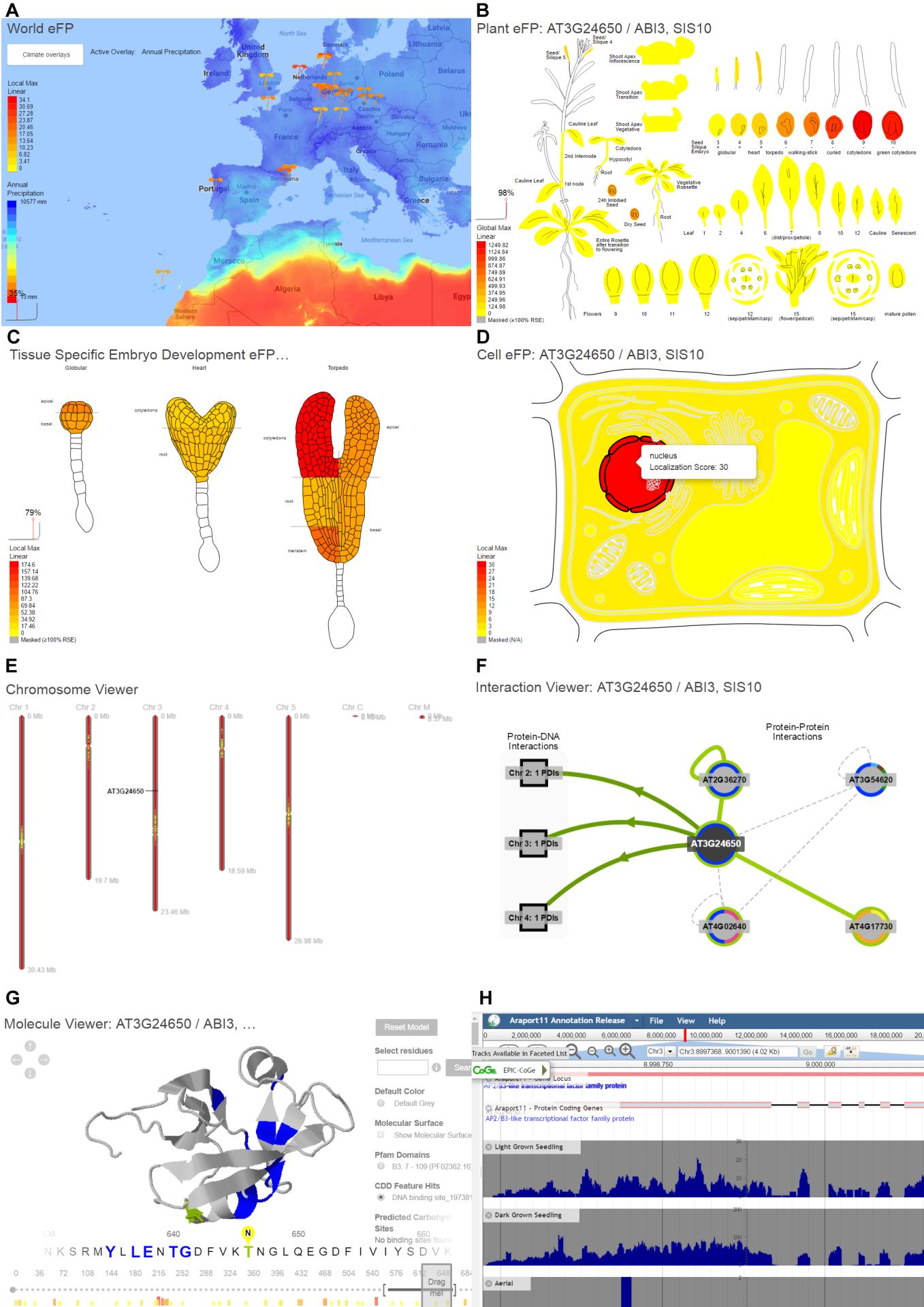
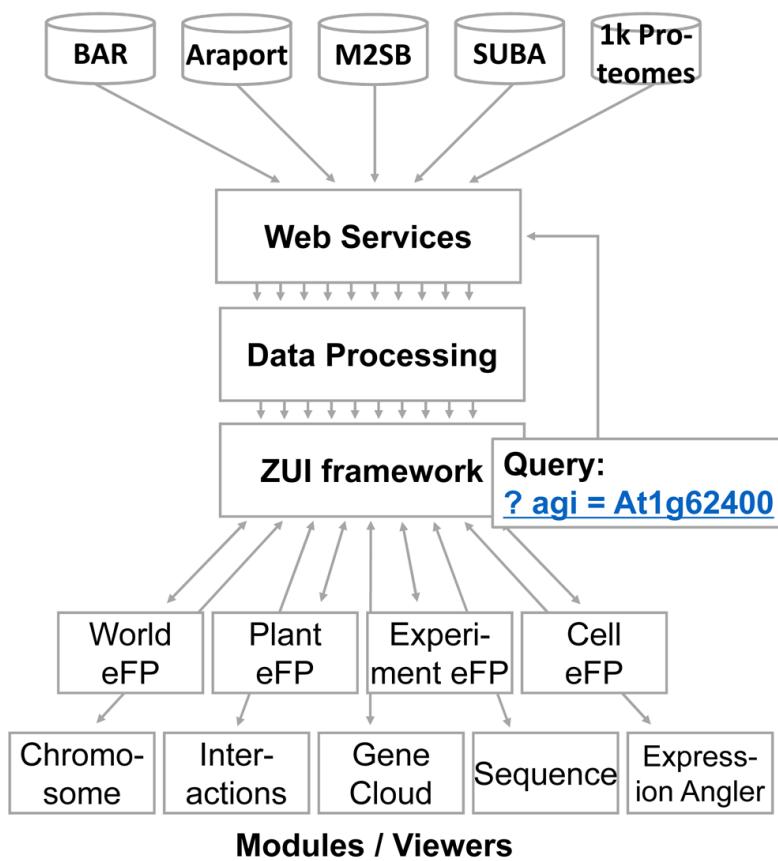


Figure 2. ePlant's module viewers (each displaying data for *ABI3*) are presented here with the intention of illustrating ePlant's hierarchy of scale. To see detail, view online version. (A) World eFP Viewer, (B) Plant eFP Viewer, (C) Tissue & Experiment eFP Viewer, (D) Subcellular eFP Viewer, (E) Chromosome Viewer, (F) Interactions Viewer, (G) Molecule Viewer, (H) Sequence Viewer. High resolution vector graphics are available for most views (see example in Supplemental File 1). ePlant outputs are freely usable via an "open" license.

A



B

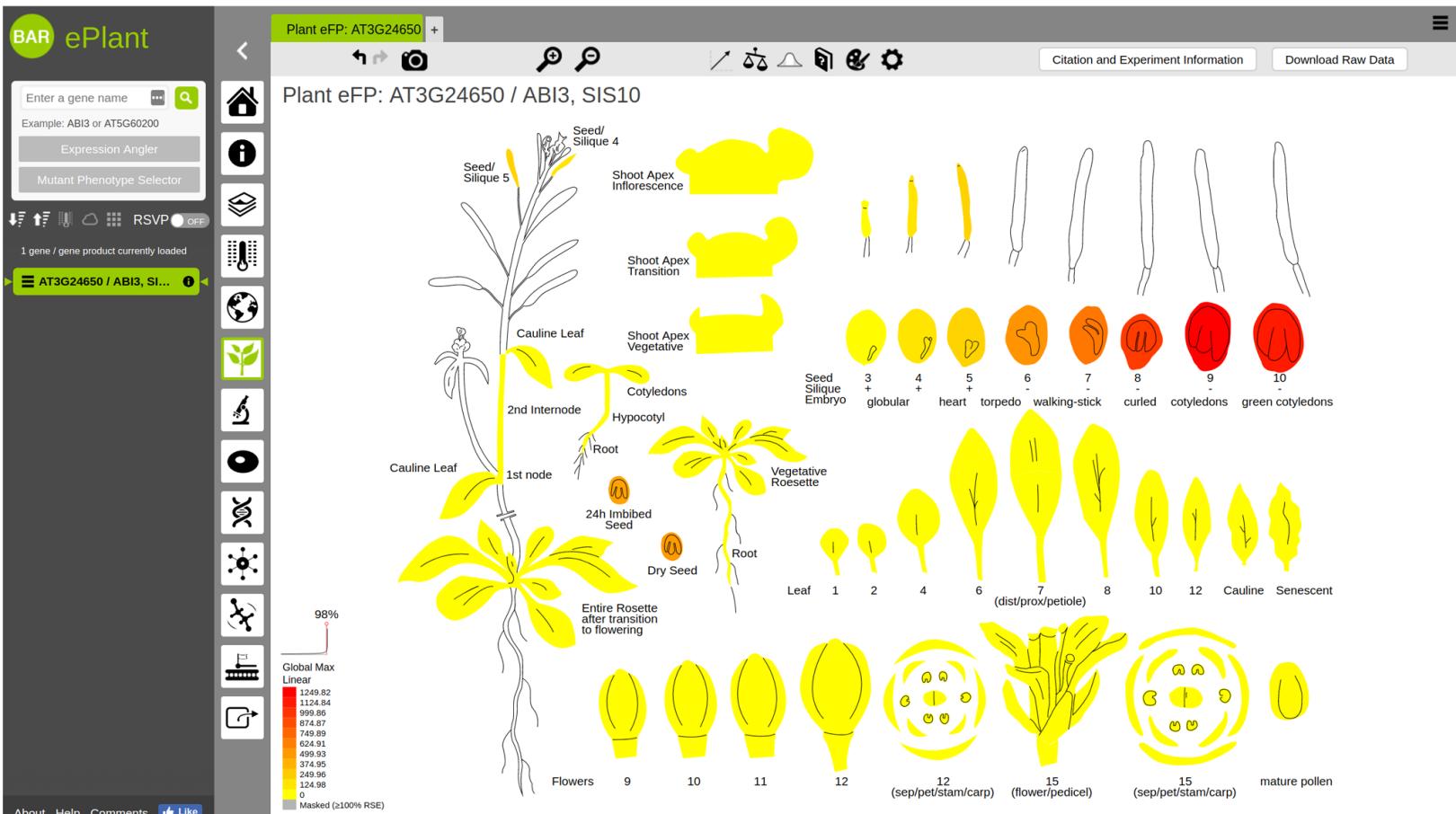


Figure 3. ePlant system design and user interface. (A) ePlant system architecture. (B) ePlant user interface showing the expression pattern of *ABI3* with the Plant eFP Viewer. See detail in online version of figure.

A**Gene Info Viewer: AT3G24650 / ...**

Location & Gene Model: Chr3: 8997370 to 9001185, Strand +

**B****Publication Viewer: AT3G24650 / ...****Publications**

Author	Year	Journal	Title	PubMed
Alonso R	2009	Plant Cell	A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of <i>Arabidopsis</i> seed maturation gene expression based on heterodimerization and protein complex formation.	19531597
Arenas-Huerto F	2000	Genes Dev	Analysis of <i>Arabidopsis</i> glucose insensitive mutants, gin5 and gin6, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar.	10950871
Arenas-Mena C	1999	Plant Mol Biol	Expression and cellular localization of Atrab28 during <i>arabidopsis</i> embryogenesis.	10412913
Arroyo A	2003	Plant Physiol	Three genes that affect sugar sensing (abscisic acid insensitive 4, abscisic acid insensitive 5, and constitutive triple response 1) are differentially regulated by glucose in <i>Arabidopsis</i> .	12970489
Bassel GW	2006	J Exp Bot	ABI3 expression ceases following, but not during, germination of tomato and <i>Arabidopsis</i> seeds.	16531465

Figure 4. ePlant views providing information from Araport. (A) ePlant Gene Information Viewer. (B) ePlant Publications Viewer.

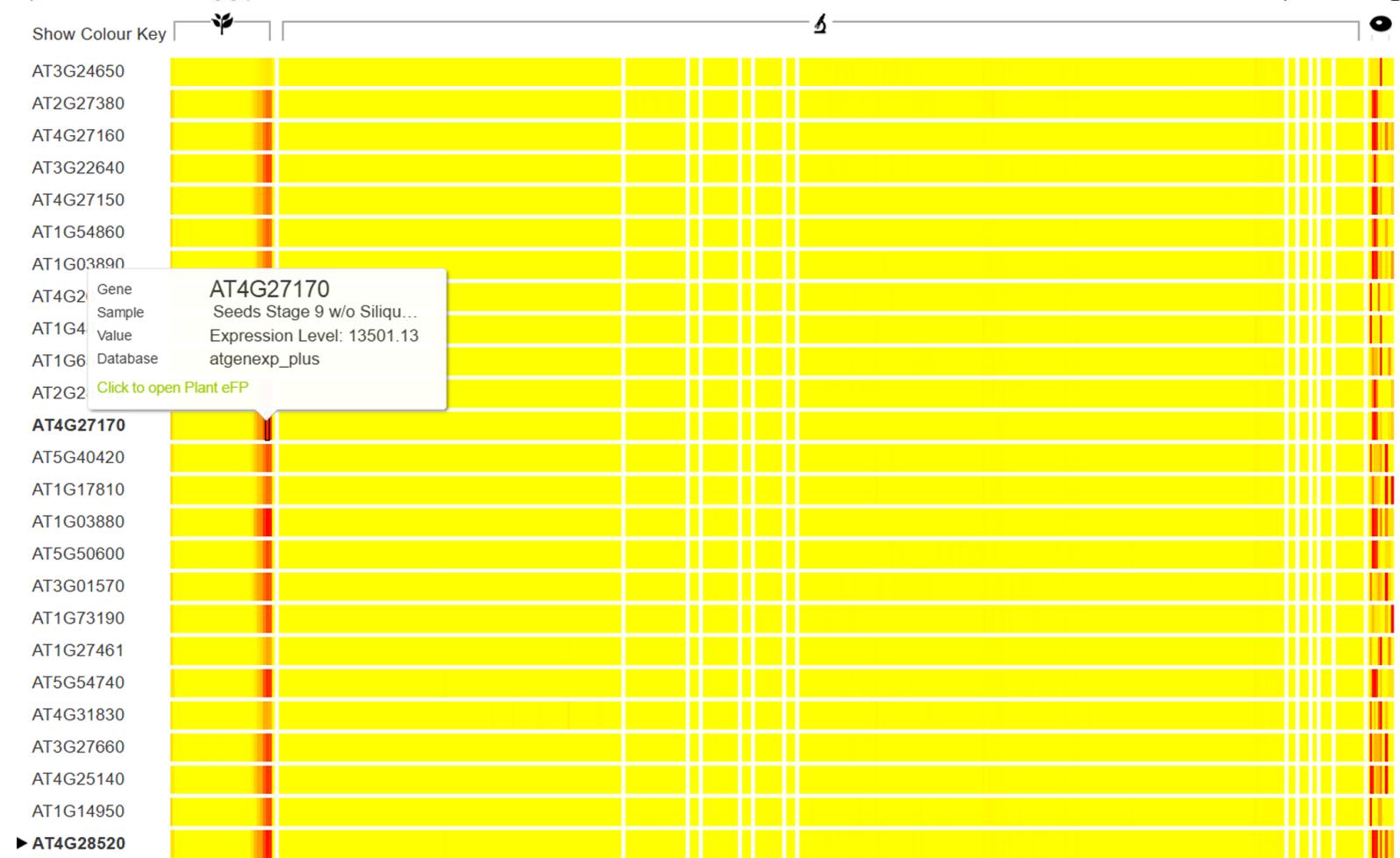
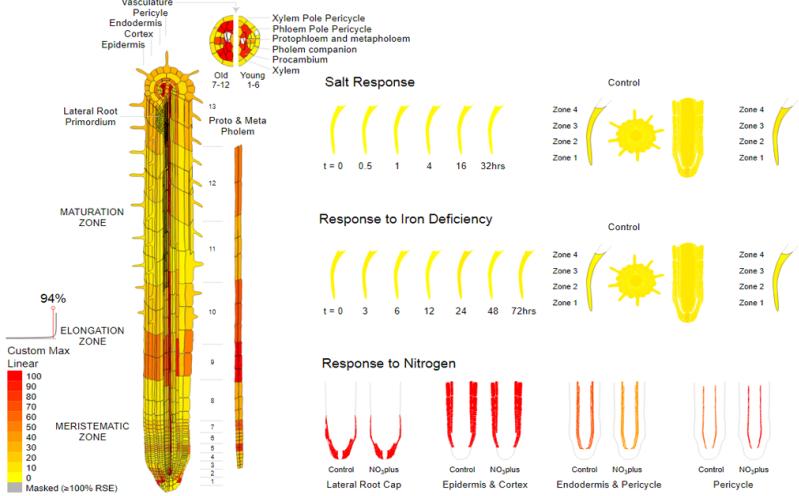


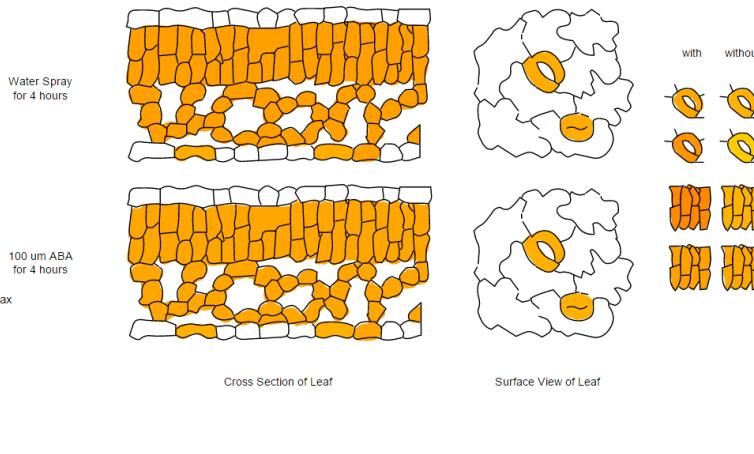
Figure 5. The Heat Map Viewer showing 350+ expression level samples for twenty-five genes identified with the Expression Angler for having similar expression patterns to *ABI3* (At3g24650). The “global” colour gradient is selected, making it easy to see the variability in the expression levels of the various genes.

A

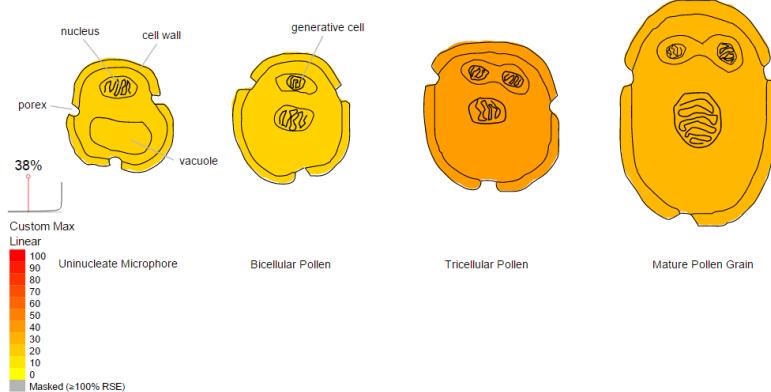
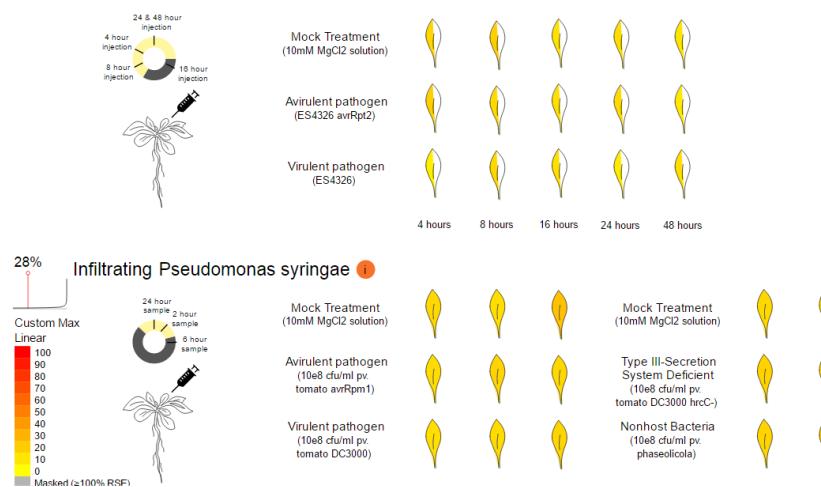
Tissue Specific Root eFP: AT3G24650 / ABI3,...

**B**

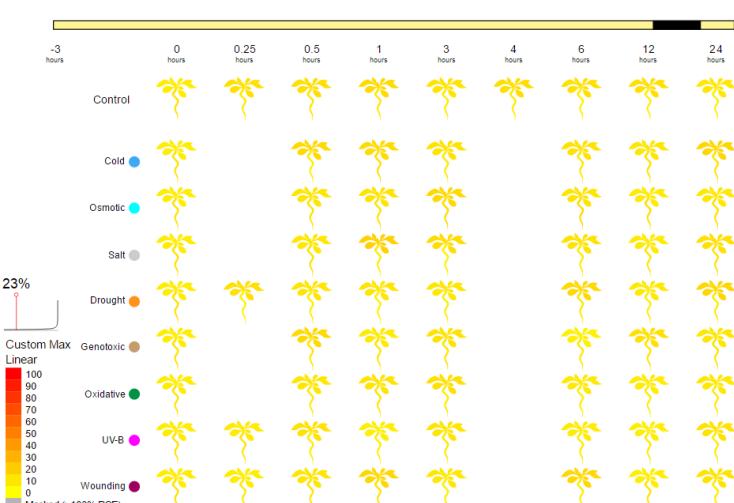
Tissue Specific Guard And Mesophyll C...

**C**

Tissue Specific Microgametogenesis e...

**D**Biotic Stress Pseudomonas syringae e...
Half Leaf Pseudomonas syringae**E**

Abiotic Stress eFP: AT3G24650 / ABI3,...

**F**

Tissue Specific Pollen Germination eFP..

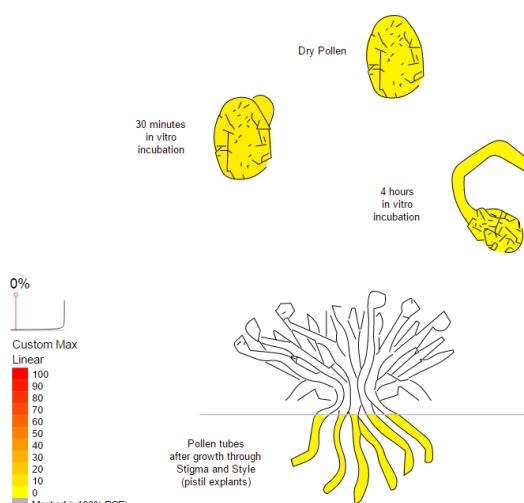


Figure 6. Six of more than twenty views from the Tissue & Experiment eFP Viewer. Each view displays expression levels for *ABI3* with the “custom” colour gradient setting, with red = 100 expression units: (A) Root, (B) Guard and Mesophyll Cells, (C) Microgametogenesis, (D) Biotic Stress: *Pseudomonas syringae*, (E) Abiotic Stress, (F) Pollen Germination. Some views are truncated for display here; see online version of figure to be able to see detail. ePlant outputs for all views may be downloaded as high-resolution vector graphic files and are freely available for use by any researcher under an “open” license.

A

Cell eFP: AT3G24650 / ABI3, SIS10

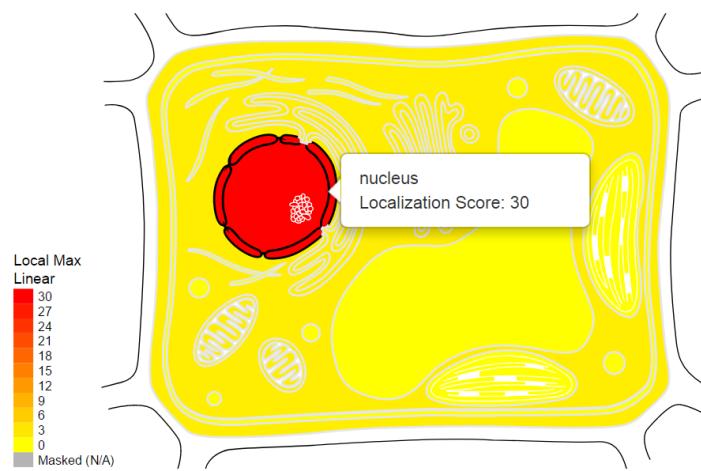
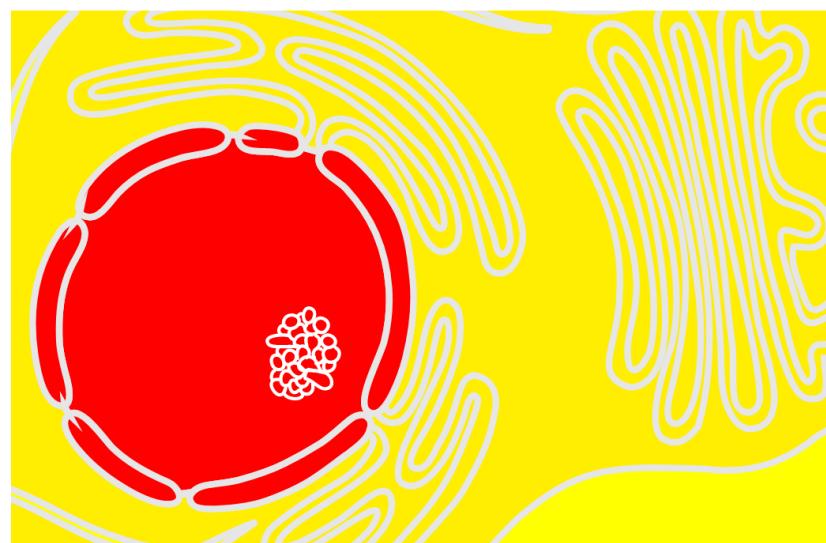
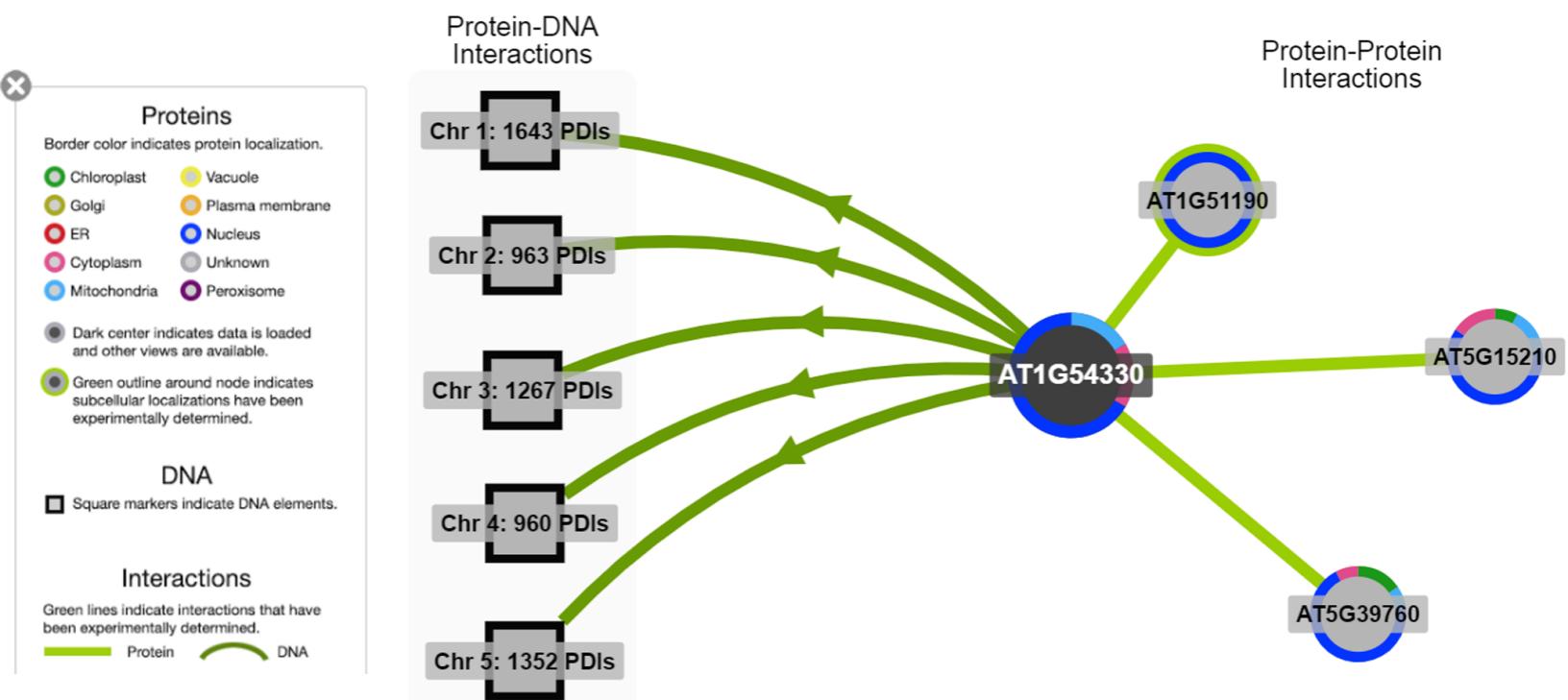
**B**

Figure 7. The Subcellular Localization eFP Viewer. (A) ABI3 is mostly localized in the nucleus. (B) An inset of a high resolution version of the same image.

A
Interaction Viewer: AT1G54330 / ANAC020, NAC020
Recursive interactions not shown



B
Molecule Viewer: AT3G24650 / ABI3, SIS10

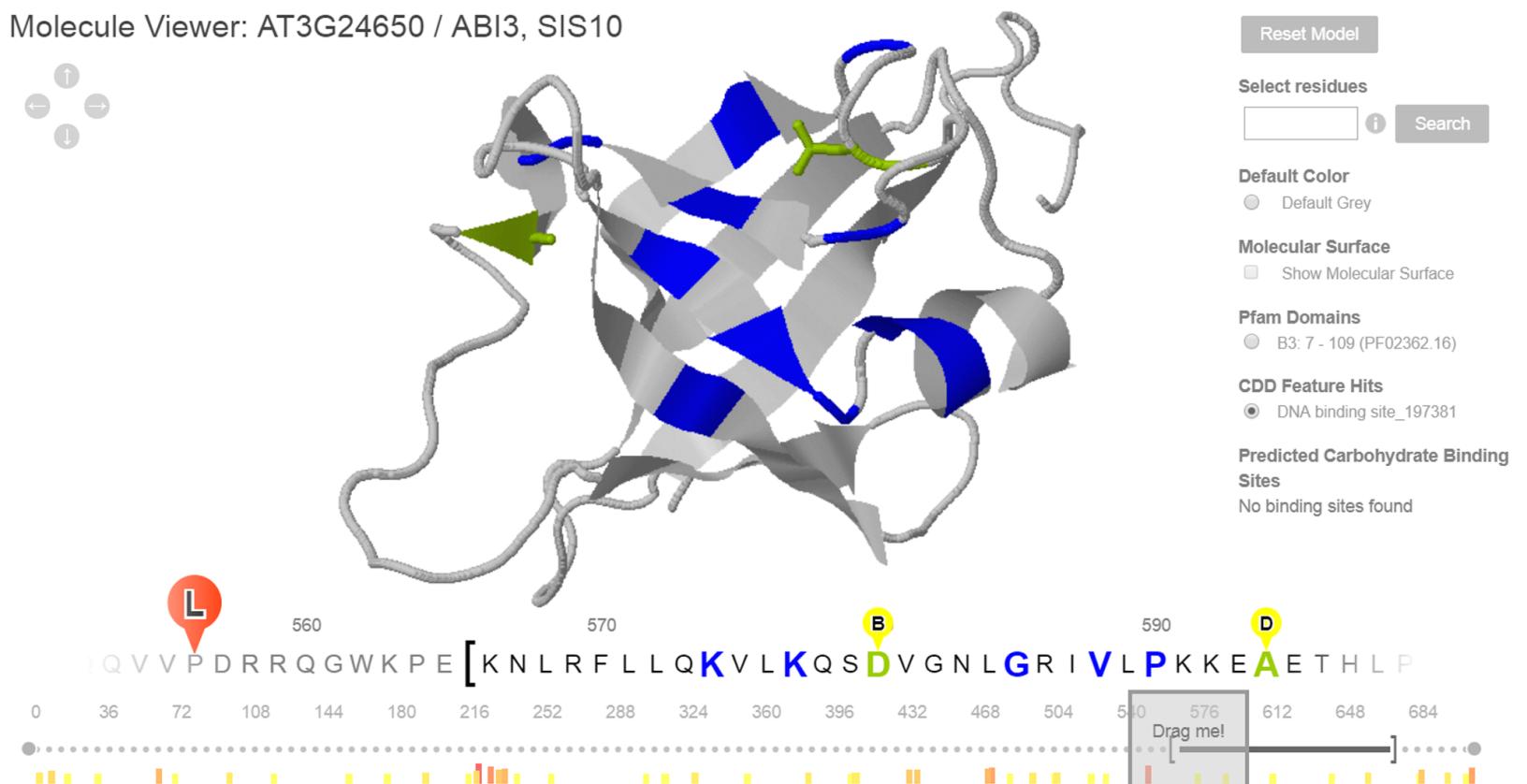


Figure 8. ePlant Interaction and Molecule Viewers.(A) Protein and DNA Interaction Viewer showing interactions for At1g54330. (B) Molecule Viewer showing the transcription factor ABI3's Phyre2-predicted partial 3D structure with its DNA binding site highlighted in blue, and two non-synonymous changes (from a web service provided by the 1001 Proteomes site) highlighted in green. Changes of higher frequency are denoted by larger, redder pins above the sequence below.

Parsed Citations

Arabidopsis Interactome Mapping Consortium (2011). Evidence for Network Evolution in an Arabidopsis Interactome Map. Science 333, 601-607.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Adams, J., Mansfield, M.J., Richard, D.J., and Doxey, A.C. (2017). Lineage-specific mutational clustering in protein structures predicts evolutionary shifts in function. Bioinformatics. doi: 10.1093/bioinformatics/btw815.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Arabidopsis Interactome Mapping Consortium (2011). Evidence for Network Evolution in an Arabidopsis Interactome Map. Science 333, 601-607. **Austin, R.S., Hiu, S., Waese, J., Ierullo, M., Pasha, A., Wang, T., Fan, J., Foong, C., Breit, R., Desveaux, D., et al. (2016). New BAR Tools for Mining Expression Data and Exploring Cis-Elements in Arabidopsis thaliana.** Plant J. 88, 490-504.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Bader, G.D., Betel, D., and Hogue, C.W.V. (2003). BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 31, 248-250.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Bederson, B.B. (2010). The Promise of Zoomable User Interfaces. In Proceedings of the 3rd International Symposium on Visual Information Communication, (New York, NY, USA: ACM). doi: 10.1145/1865841.1865844.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis Information Resource: Making and mining the "gold standard" annotated reference plant genome. Genesis 53, 474-485.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A Gene Expression Map of the Arabidopsis Root. Science 302, 1956-1960.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Böhmer, M., and Schroeder, J.I. (2011). Quantitative transcriptomic analysis of abscisic acid-induced and reactive oxygen species-dependent expression changes and proteomic profiling in Arabidopsis suspension cells. Plant J. 67, 105-118.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Casson, S., Spencer, M., Walker, K., and Lindsey, K. (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. Plant J. 42, 111-123.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S.Y. (2014). Genomic signatures of specialized metabolism in plants. Science 344, 510-513.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Chatr-Aryamontri, A., Breitkreutz, B.-J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., et al. (2015). The BioGRID interaction database: 2015 update. Nucleic Acids Res. 43, D470-478.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Couldridge, C., Newbury, H.J., Ford-Lloyd, B., Bale, J., and Pritchard, J. (2007). Exploring plant responses to aphid feeding using a full Arabidopsis microarray reveals a small number of genes with significantly altered expression. Bull. Entomol. Res. 97, 523-532.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222-D230.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Fucile, G., Di Biase, D., Nahal, H., La, G., Khodabandeh, S., Chen, Y., Easley, K., Christendat, D., Kelley, L., Provart, N.J. (2011) ePlant and the 3D Data Display Initiative: Integrative Systems Biology on the World Wide Web. *PLoS ONE* 6: e15237.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M. (2007). A predicted interactome for *Arabidopsis*. *Plant Physiol.* 145, 317-329.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Gilding, E.K., and Marks, M.D. (2010). Analysis of purified glabra3-shapeshifter trichomes reveals a role for NOECK in regulating early trichome morphogenic events. *Plant J.* 64, 304-317.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Godz, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., Ogawa, M., Yamauchi, Y., Preston, J., Aoki, K., et al. (2008). The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* 55, 526-542.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., and Sussman, J.L. (2013). JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia. *Isr. J. Chem.* 53, 207-216.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Harris, I., Jones, P. d., Osborn, T. j., and Lister, D. h. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int. J. Climatol.* 34, 623-642.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Healey, C.G., and Enns, J.T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 1170-1188.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Helt, G.A., Nicol, J.W., Erwin, E., Blossom, E., Blanchard, S.G., Chervitz, S.A., Harmon, C., and Lorraine, A.E. (2009). Genoviz Software Development Kit: Java tool kit for building genomics visualization applications. *BMC Bioinformatics* 10, 266.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Highsmith, J., and Cockburn, A (2001). Agile software development: the business of innovation. *Computer* 34, 120-127.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Hony, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* 5, R85.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Hudson, W. (2014). Agile UX and UCD. In CHI '14 Extended Abstracts on Human Factors in Computing Systems, (New York, NY, USA: ACM), pp. 1009-1010.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

International Arabidopsis Informatics Consortium (2012). Taking the Next Step: Building an *Arabidopsis* Information Portal. *Plant Cell* 24, 2248-2256.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Joshi, H.J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., Schulze, W.X., Sun, Q., Wijk, K.J. van, Egelhofer, V., et al. (2011). MASCP Gator: An Aggregation Portal for the Visualization of Arabidopsis Proteomics Data. *Plant Physiol.* 155, 259-270.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., Bono, B. de, Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., et al. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* 33, D428-D432.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845-858.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 50, 347-363.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Klein, C., and Bederson, B.B. (2005). Benefits of Animated Scrolling. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, (New York, NY, USA: ACM), pp. 1965-1968.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Krishnakumar, V., Hanlon, M.R., Contrino, S., Ferlanti, E.S., Karamycheva, S., Kim, M., Rosen, B.D., Cheng, C.-Y., Moreira, W., Mock, S.A., et al. (2014). Araport: the Arabidopsis Information Portal. *Nucleic Acids Res.* 43, D1003-D1009.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lalonde, S., Sero, A., Pratelli, R., Pilot, G., Chen, J., Sardi, M.I., Parsa, S.A., Kim, D.-Y., Acharya, B.R., Stein, E.V., et al. (2010). A membrane protein/signaling protein interaction network for Arabidopsis version AMPv2. *Front. Physiol.* doi: 10.3389/fphys.2010.00024.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., and Weigel, D. (2005). Diversity of Flowering Responses in Wild *Arabidopsis thaliana* Strains. *PLoS Genet.* 1, e6.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lloyd, J., and Meinke, D. (2012). A Comprehensive Dataset of Genes with a Loss-of-Function Mutant Phenotype in Arabidopsis. *Plant Physiol.* 158, 1115-1129.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q., and Bader, G.D. (2010). Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26, 2347-2348.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222-226.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Marks, M.D., Wenger, J.P., Gilding, E., Jilk, R., and Dixon, R.A. (2009). Transcriptome Analysis of Arabidopsis Wild-Type and gl3-sst sim Trichomes Identifies Four Additional Genes Required for Trichome Development. *Mol. Plant* 2, 803-822.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* 41, 697-709.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nawy, T., Lee, J.-Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., and Benfey, P.N. (2005). Transcriptional Profile of the *Arabidopsis* Root Quiescent Center. *Plant Cell* 17, 1908-1925.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nielsen, J. (1993). **Usability Engineering** (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.).

Nielson, J., and Landauer, T. A mathematical model of the finding of usability problems. *Proc. ACM INTERCHI93 Conf. Amst. Neth.* 24-29 April 1993, 206-213.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 165, 1280-1292.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pandey, S., Wang, R.-S., Wilson, L., Li, S., Zhao, Z., Gookin, T.E., Assmann, S.M., and Albert, R. (2010). Boolean modeling of transcriptome data reveals novel modes of heterotrimeric G-protein action. *Mol. Syst. Biol.* 6, 372.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pillitteri, L.J., Peterson, K.M., Horst, R.J., and Torii, K.U. (2011). Molecular Profiling of Stomatal Meristemoids Reveals New Component of Asymmetric Cell Division and Commonalities among Stem Cell Populations in *Arabidopsis*. *Plant Cell* 23, 3260-3275.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Popescu, S.C., Snyder, M., and Dinesh-Kumar, S.P. (2007). Arabidopsis Protein Microarrays for the High-Throughput Identification of Protein-Protein Interactions. *Plant Signal. Behav.* 2, 416-420.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Popescu, S.C., Popescu, G.V., Bachan, S., Zhang, Z., Gerstein, M., Snyder, M., and Dinesh-Kumar, S.P. (2009). MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes Dev.* 23, 80-92.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A., and Palanivelu, R. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.* 5, e1000621.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501-506.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Schneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, (IEEE), pp. 336-343.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., and Holmes, I.H. (2009). JBrowse: A next-generation genome browser. *Genome Res.* 19, 1630-1638.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Speier, C., Vessey, I., and Valacich, J.S. (2003). The Effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision-Making Performance. Decis. Sci. 34, 771-797.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Spence, R. (2002). Rapid, Serial and Visual: A Presentation Technique with Potential. Inf. Vis. 1, 13-19.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005). Cuticular Lipid Composition, Surface Structure, and Gene Expression in *Arabidopsis* Stem Epidermis. Plant Physiol. 139, 1649-1665.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Swanson, R., Clark, T., and Preuss, D. (2005). Expression profiling of *Arabidopsis* stigma tissue identifies stigma-specific genes. Sex. Plant Reprod. 18, 163-171.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Tanz, S.K., Castleden, I., Hooper, C.M., Vacher, M., Small, I., and Millar, H.A. (2013). SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. Nucleic Acids Res. 41, D1185-1191.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Taylor-Teeple, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., et al. (2015). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. Nature 517, 571-575.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J. (2005). The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. Plant J. 43, 153-163.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Tufte, E.R. (1990). Envisioning Information (Graphics Press, Cheshire, CT).

Tufte, E.R., and Graves-Morris, P.R. (1983). The visual display of quantitative information (Graphics Press, Cheshire, CT).

Waese, J., Stuerzlinger, W., and Provart, N.J. (2016). An Evaluation of Interaction Methods for Controlling RSVP Displays in Visual Search Tasks. In Big Data Visual Analytics (BDVA), 2016, (Sydney: IEEE Xplore). doi: 10.1109/BDVA.2016.7787041.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D., Fu, X.-D., and Dong, X. (2011). Timing of plant immune responses by a central circadian regulator. Nature 470, 110-114.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Ware, C. (2012). Information Visualization: Perception for Design (Elsevier).

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE 2, e718.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Wisecaver, J.H., Borowsky, A.T., Tzin, V., Jander, G., Kliebenstein, D.J., and Rokas, A. (2017). A Global Co-expression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell tpc.00009.2017.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V. (2009). Gene expression map of the *Arabidopsis* shoot apical meristem stem cell niche. Proc. Natl. Acad. Sci. U. S. A. 106, 4941-4946.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Yang, Y., Costa, A., Leonhardt, N., Siegel, R.S., and Schroeder, J.I. (2008). Isolation of a strong *Arabidopsis* guard cell promoter and its potential as a research tool. *Plant Methods* 4, 6.

Pubmed: [Author and Title](#)

CrossRef: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology

Jamie Waese, Jim Fan, Asher Pasha, Hans Yu, Geoffrey Fucile, Ruian Shi, Matthew Cumming, Lawrence Kelley, Michael Sternberg, Vivek Krishnakumar, Erik Ferlanti, Jason Miller, Chris Town, Wolfgang Stuerzlinger and Nicholas J. Provert

Plant Cell; originally published online August 14, 2017;
DOI 10.1105/tpc.17.00073

This information is current as of September 7, 2017

Supplemental Data

</content/suppl/2017/08/23/tpc.17.00073.DC1.html>

Permissions

https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X

eTOCs

Sign up for eTOCs at:

<http://www.plantcell.org/cgi/alerts/ctmain>

CiteTrack Alerts

Sign up for CiteTrack Alerts at:

<http://www.plantcell.org/cgi/alerts/ctmain>

Subscription Information

Subscription Information for *The Plant Cell* and *Plant Physiology* is available at:

<http://www.aspb.org/publications/subscriptions.cfm>