

# SCIENTIFIC REPORTS



OPEN

## Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination

Daniel Bellieny-Rabelo<sup>1</sup>, Eduardo Alves Gamosa de Oliveira<sup>1</sup>, Elane da Silva Ribeiro<sup>1,2</sup>, Evenilton Pessoa Costa<sup>1,3</sup>, Antônia Elenir Amâncio Oliveira<sup>1</sup> & Thiago Motta Venancio<sup>1</sup>

Received: 12 August 2016

Accepted: 10 October 2016

Published: 08 November 2016

Soybean (*Glycine max*) is a major legume crop worldwide, providing a critical source of protein and oil. The release of the soybean genome fuelled several transcriptome projects comprising multiple developmental stages and environmental conditions. Nevertheless, the global transcriptional patterns of embryonic axes during germination remain unknown. Here we report the analysis of ~1.58 billion RNA-Seq reads from soybean embryonic axes at five germination stages. Our results support the early activation of processes that are critical for germination, such as glycolysis, Krebs cycle and cell wall remodelling. Strikingly, only 3 hours after imbibition there is a preferential up-regulation of protein kinases and transcription factors, particularly from the LOB domain family, implying that transcriptional and post-transcriptional regulation play major roles early after imbibition. Lipid mobilization and glyoxylate pathways are also transcriptionally active in the embryonic axes, indicating that the local catabolism of oil reserves in the embryonic axes contributes to energy production during germination. We also present evidence supporting abscisic acid inactivation and the up-regulation of gibberellin, ethylene and brassinosteroid pathways. Further, there is a remarkable differential activation of paralogous genes in these hormone signalling pathways. Taken together, our results provide insights on the regulation and biochemistry of soybean germination.

Soybean [*Glycine max* (L.) Merr.] is the most economically important legume crop and a key factor of global food security due to its rich oil and protein content<sup>1</sup>. *G. max* belongs to the family Fabaceae, which is the third largest angiosperm family, comprising nearly 19,500 species<sup>2</sup>. Over the past 10 years, several legume species had their genomes sequenced, such as soybean<sup>3</sup>, common bean (*Phaseolus vulgaris*)<sup>4</sup>, chickpea (*Cicer arietinum*)<sup>5</sup> and the diploid ancestors of the cultivated peanut<sup>6</sup>.

During maturation, cotyledonary cells accumulate massive amounts of reserves and other macromolecules that will be necessary during germination<sup>7</sup>. Conversely, the embryonic axis is not a primary source of reserves; instead, it harbours organs' primordia, namely the shoot apex, hypocotyl and radicle<sup>8</sup>. Nevertheless, the embryonic axes of some species can grow and develop seedlings in the absence of cotyledons<sup>9</sup>, mainly because of stored maternal mRNAs, minor energetic reserves and enzymes that allow early metabolic resumption and growth upon imbibition<sup>10</sup>. During development, seed coat formation and progressive desiccation result in the cessation of metabolic activity. The period between the completion of seed development and germination is called after-ripening, a stage that results in increased sensitivity to factors that promote germination and decreased sensitivity to those that repress it<sup>9</sup>. When seeds are non-dormant and favourable environment conditions are met (e.g. when water and oxygen reach the embryo), germination starts and embryonic tissues resume growth<sup>10</sup>. Germination can be divided in early and late phases; the former comprises imbibition, DNA and organelle repair, metabolism reactivation, endosperm weakening and embryo elongation; whereas the latter involves radicle protrusion and rupture of the endosperm and seed coat<sup>8,10</sup>. Maintenance and release of seed dormancy are promoted by abscisic

<sup>1</sup>Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, Brazil. <sup>2</sup>Núcleo em Ecologia e Desenvolvimento Sócio-Ambiental de Macaé (NUPEM), Campus UFRJ Macaé, Macaé, Brazil. <sup>3</sup>Unidade de Experimentação Animal, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, Brazil. Correspondence and requests for materials should be addressed to T.M.V. (email: thiago.venancio@gmail.com)

Stage	Unmapped reads	Multiple matches	Uniquely mapped	Total mapped reads
Dry	21,888,018	85,630,708	157,064,044	242,694,752
3-HAI	16,427,943	71,875,156	182,185,569	254,060,725
6-HAI	19,798,969	87,809,774	211,697,839	299,507,613
12-HAI	17,874,585	76,704,465	200,731,192	277,435,657
24-HAI	24,183,233	97,381,607	275,170,244	372,551,851

**Table 1. Alignment of RNA-Seq reads on the soybean genome.** \*HAI, hours after imbibition.

acid (ABA) and gibberellic acid (GA), respectively<sup>8,10,11</sup>; further, key genes involved in the GA/ABA interplay have been characterized [e.g. 9-*cis* epoxy-carotenoid dioxygenase (*NCED*), gibberellin 2-oxidase (*GA2ox*), gibberellin insensitive dwarf 1 (*GIDI*) and phytochrome-interacting factors (PIF's)], which helped elucidate underlying aspects of ABA and GA regulation<sup>12–14</sup>.

Several large-scale transcriptomic studies were conducted for a variety of dicot species, such as *Arabidopsis thaliana*<sup>15</sup> and chickpea<sup>16</sup>. Further, soybean transcriptomic studies allowed the discovery of important genes involved in the determination of floral organs and maintenance of shoot apical meristem<sup>17</sup>; seed filling<sup>18</sup>; response against drought stress<sup>19</sup> and regulation of seed development<sup>20</sup>. Others found stage-specific alternative splicing isoforms of important genes<sup>21</sup>. However, the transcriptomic changes of soybean embryonic axes during germination remain unexplored, in spite of the essential nature of this stage in plant growth. Here we report the analysis of ~1.58 billion RNA-Seq reads from soybean embryonic axes in 5 seed germination time points. Among our main results are: i) a strong preferential activation of transcription factors (TFs) and protein kinases in the onset of germination; ii) an important carbohydrate partitioning system underlying cell wall metabolism and ATP production; iii) an early activation of primary catabolic pathways; iv) a strong up-regulation of genes involved in hormone biosynthesis, catabolism and signalling; iv) transcriptional divergence and potential sub-functionalization of paralogous genes with critical roles in hormone signalling, (e.g. *GIDI*, *DELLA* and *CYP707A*).

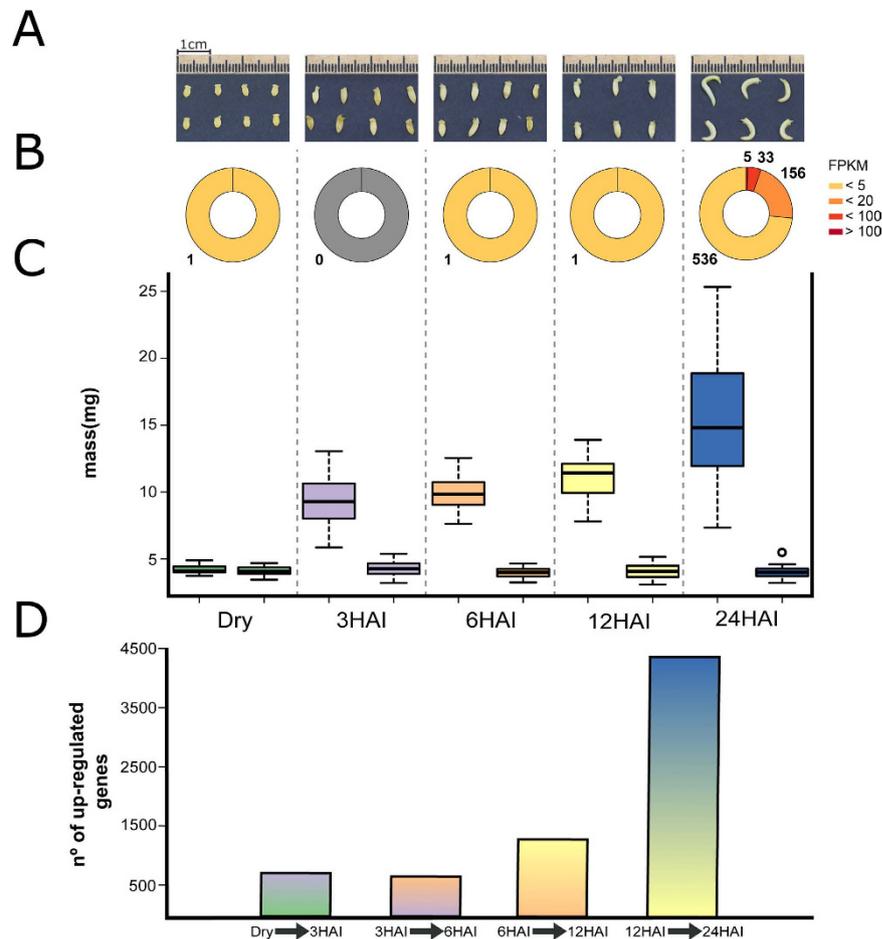
## Results and Discussion

**Transcriptome sequencing of embryonic axes during soybean germination.** Aiming to understand the transcriptional landscape of soybean embryonic axes during germination, we sequenced mRNA libraries of embryonic axes at 5 different germination time points: dry (0 hours after imbibition, HAI), 3-HAI, 6-HAI, 12-HAI and 24-HAI. A total of 242 to 372 million paired-end reads were sequenced per stage (Table 1). In terms of depth, our dataset (~1.58 billion reads) is ~2.4 times greater than that produced by a recent *G. max* RNA-Seq study<sup>21</sup>. The data reported here are also of good quality, as 91.8% (~1.45 billion) of the reads were successfully mapped on the soybean reference genome, out of which ~1.03 billion (~71.0%) were uniquely mapped (Table 1). We found detectable transcription (FPKM > 1; see methods for details) for 33,305 (61.48%) genes predicted in the *G. max* reference genome (Table S1).

Our results support the notion that the mRNA populations of early germinating seeds comprise two distinct subpopulations: transcripts that are important at late maturation (e.g. Late Embryogenesis Abundant proteins, LEAs) and; transcripts that will be required for early germination phases (Table S1; Figures S1 and S2)<sup>8,10,22</sup>. In an initial analysis, we found that 99.6% of the preferentially expressed genes (730 out of 733; see methods for details) are remarkably activated at 24-HAI (Fig. 1A; Table S2). Although it would be intuitive to associate this dramatic shift with major morphological transitions, no significant dry weight changes were observed (Fig. 1), implying that the embryonic axis growth during germination is mainly driven by water intake and cell expansion<sup>8</sup>. Photosynthesis (GO:0015979) and response to oxidative stress (GO:0006979) genes are remarkably present in this 24-HAI preferentially expressed gene set. We hypothesize that this ~700-fold enrichment in preferentially expressed genes constitutes a preparation phase for seedling establishment and autotrophic seedling growth (discussed later).

**Genes involved in cell-wall homeostasis, protein kinases and TFs are up-regulated early during germination.** We detected 730, 674, 1307 and 4394 up-regulated genes in the dry to 3-HAI, 3-HAI to 6-HAI, 6-HAI to 12-HAI and 12-HAI to 24-HAI intervals, respectively. To uncover major trends in these gene sets, we performed a Gene Ontology (GO) enrichment analysis using stringent statistical criteria (Table S3). This analysis revealed an early (i.e. at 3-HAI) significant up-regulation of cell wall genes (GO:0005618; 14 genes), particularly xyloglucan endotransglucosylases (GO:0016762; 10 genes; e.g. *Glyma11g36730*, 77-fold up-regulated), which promote cellular expansion by remodelling cellulose microfibrils. Many other cell wall genes were up-regulated at 24-HAI (43 genes) (Table S3), although this second gene set includes 15 cellulose synthases, which are critical for cell wall homeostasis. There was also a clear enrichment of genes annotated with the general term GO:0005975 (i.e. “carbohydrate metabolic process”) at 24-HAI (Table S3), encompassing several gene families involved in cell wall organization or defence mechanisms (e.g. beta-glucosidases and pectin lyases).

The only GO term significantly enriched at all stages was GO:0016705, which comprises a variety of oxidoreductase families, such as: 1) ACC oxidases ACO1 and ACO4, which catalyse the last step in ethylene biosynthesis. Interestingly, 5 out of 6 differentially expressed ACO4 genes reach high expression levels (FPKM > 10) only after 6-HAI, indicating that ethylene is more relevant later in germination<sup>23</sup>; 2) 47 cytochrome P450 genes (i.e. CYP), which are mainly induced between 12 and 24-HAI (33 genes); 3) GA oxidases (8 genes) and; 4) 28 oxoglutarate/iron-dependent dioxygenases. Further, we found a related GO term, GO:0055114



**Figure 1.** (A) Photographs of soybean embryonic axes at 0- (dry), 3-, 6-, 12- and 24-HAI; (B) Number of preferentially expresses genes and their transcriptional levels (in FPKM); (C) Distribution of fresh and dry weight of embryonic axes (20 seeds per condition); (D) Number of up-regulated genes between sequential germination time points.

(oxidation-reduction process), enriched among genes induced at 6-, 12- and 24-HAI; this group includes aldehyde dehydrogenases, FAD-binding Berberine family proteins, lipoxigenases, peroxidases, Sks cell wall oxidoreductases and rhamnose synthases. Although structurally distinct, several of these families have been linked with cell wall metabolism and homeostasis<sup>24–26</sup>.

We also found a remarkable enrichment of several GO terms related to photosynthesis (e.g. GO:0009579, thylakoid; GO:0015995, chlorophyll biosynthetic process), which we unified under the GO term “photosynthesis” (GO:0015979). This large group (77 genes induced at 12- and/or 24-HAI) comprise multiple photosystem I and II subunits, including genes encoding proteins of the light-harvesting complex and chlorophyll binding proteins. Many of these genes are also preferentially expressed at 24-HAI (discussed above) and their expression late in germination probably integrate a preparation phase for the transition to autotrophic growth.

The GO enrichment analyses also pointed us to two remarkable aspects of the soybean germination transcriptome: the preferential up-regulation of genes encoding protein kinases (GO:0004672) and transcriptional regulators (GO:0030528) at 3-HAI (Table S3). Although most of the up-regulated protein kinases are transcribed at moderate levels (42 genes; FPKM < 10), 7 were highly transcribed (FPKM > 10): *Glyma09g32390* (a proline extensin-like receptor kinase) and *Glyma07g38510* (a mitogen-activated kinase) are not well characterized and their early up-regulation is followed by a subsequent down-regulation, suggesting specific roles in the germination onset. *Glyma06g15270* is a brassinosteroid (BR) receptor (homolog of ATBRI1/AT4G39400) that is 4-fold up-regulated at 3-HAI. BR signalling via ATBRI1 has been shown to promote germination by counteracting ABA effects<sup>27</sup>. Further, *Glyma06g15270* is up-regulated in response to exogenous BR and is able to restore the *Arabidopsis* dwarf mutant *bri1-5* phenotype<sup>28</sup>. Strikingly, *Glyma12g22660*, the homolog of the *Arabidopsis* *THESEUS1*, is also up-regulated at 3-HAI. *THESEUS1* is induced by BR and is required for optimal cell elongation<sup>29</sup>. Further, *THESEUS1* attenuates the inhibition of hypocotyl growth in some cell-wall mutants<sup>30</sup> and integrates a cell-wall surveillance system in expanding cells<sup>30</sup>. Interestingly, *Glyma12g22660* is the only one of these 7 highly up-regulated protein kinases that is also significantly induced between 12 and 24-HAI. This bimodal pattern of activation at 3- and 24-HAI coincides with the second water uptake phase and with a massive up-regulation of cell wall genes (Table S3). Therefore, we hypothesize that BR signalling via *BRI1* and *THESEUS1* constitutes an important system that regulates cell expansion during germination. Another kinase that is highly

induced at 3-HAI is *Glyma10g33240*, an ethylene receptor whose homolog functions in *Arabidopsis* seedlings<sup>31</sup>. The sixth up-regulated kinase is *Glyma14g39290*, a homolog of the *Arabidopsis* transmembrane kinase *TMK1* (*AT1G66150*), reported to be transcriptionally induced by GA<sup>32</sup>. Finally, we found an apparently unexpected up-regulation of *Glyma07g10695*, a homolog of the *Arabidopsis* *AT1G18390*, which is positively involved in ABA signalling<sup>33</sup>. Overall, our results suggest that signalling cascades involving protein phosphorylation play critical roles in the germination onset.

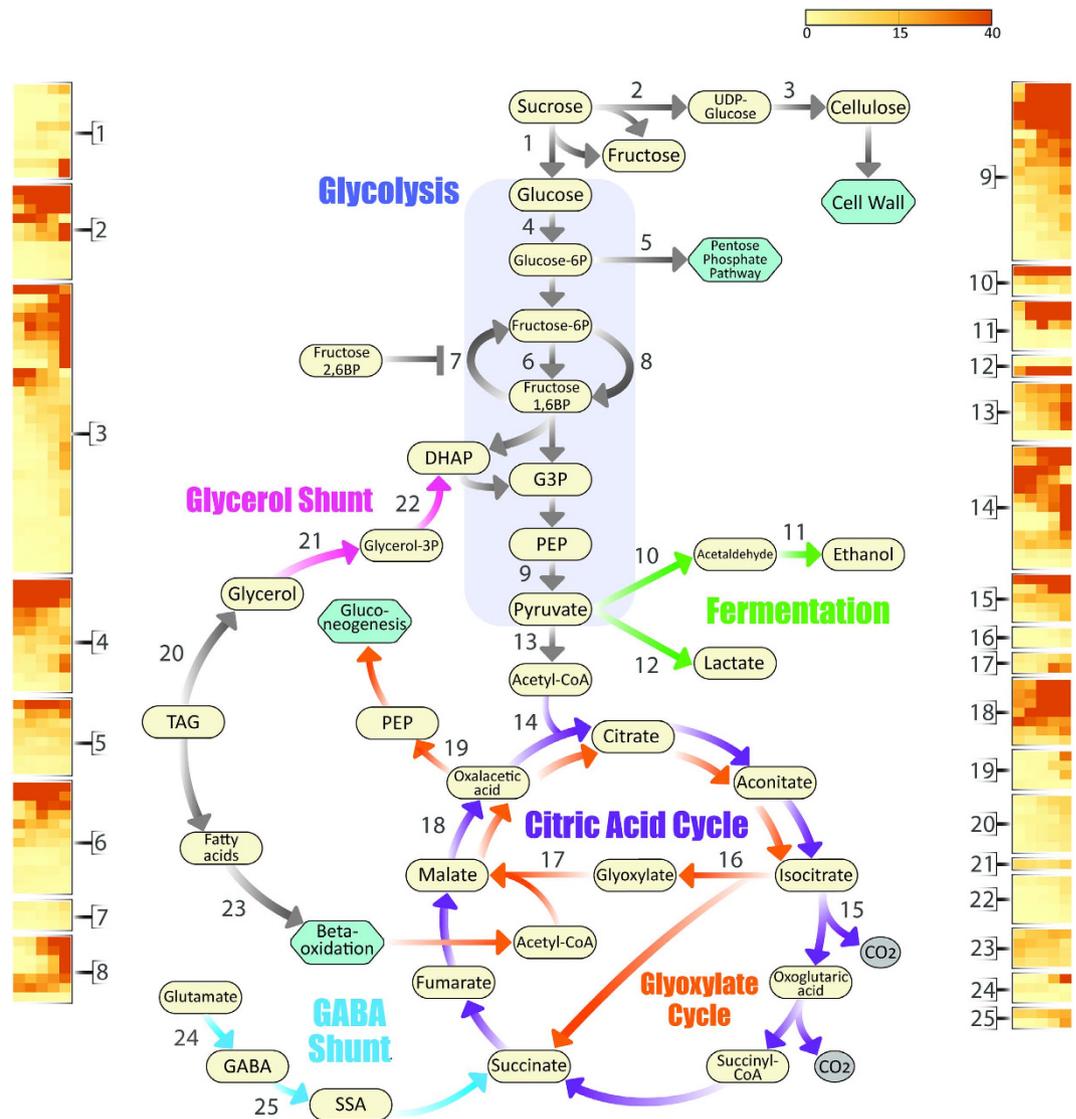
Perhaps the most remarkable aspect of the genes up-regulated at 3-HAI is the massive presence of DNA-dependent transcriptional regulators (GO:0006355) (Table S3), most of which annotated as DNA-binding TFs in PlantTFDB<sup>34</sup>. Since PlantTFDB is a public repository specialized on TFs, we analysed the differential expression of TFs as per this database classifications. We detected transcription of 53.8% of the soybean TFs (2000/3717) during germination, out of which 432 were significantly up-regulated in at least one stage. Notably, 3-HAI was the only stage with significant over-representation of TFs (81 genes;  $p$ -value =  $8 \times 10^{-8}$ ; Fisher Exact Test), corroborating the GO enrichment results. Overall, ~11% of the genes up-regulated at 3-HAI are TFs, which allowed us to uncover other remarkable trends: 1) the striking over-representation of LOB domain genes induced early during germination<sup>35</sup>; 9 out of 81 TFs up-regulated at 3-HAI are from the LOB family, which has a total of 36 members expressed in at least one germination time point. The transcriptional fold-changes of the LOB genes are also very high; for example, *Glyma07g10180* and *Glyma13g07590* have low expression levels (<2 FPKM) in dry seeds and are greatly up-regulated by 3-HAI (>96 FPKM). These observations are coherent with our previous discussion on BR signalling, as BRs induce the expression of LOB TFs, which in turn repress the accumulation of BRs via up-regulation of the BR-inactivating enzyme *BAS1*<sup>36</sup>. Therefore, LOB TFs are part of a feedback loop involved in BR homeostasis<sup>37</sup>. Nevertheless, because *BAS1* genes were transcribed at low levels during soybean germination, this feedback loop is either inactive or involves other genes at this phase. Further, LOB TFs promote the transcription of an expansin and probably other cell wall-loosening genes<sup>38</sup>. Thus, we hypothesize that LOB TFs regulate cell-wall homeostasis during germination; 2) the 3.2-fold up-regulation of *Glyma02g08835*, an ethylene response factor gene. *Glyma02g08835* is highly expressed in dry seeds (FPKM = 72.9), further up-regulated at 3-HAI (FPKM = 237) and sharply down-regulated at 24-HAI (FPKM = 7.55), suggesting that it is specifically required in the onset of germination. Interestingly, this transient up-regulation seems to be a general trend of ERF TFs; 9 out of the 13 ERFs that are up-regulated at 3-HAI have their levels reduced by at least 2-fold by 24-HAI; 3) Finally, we have also noticed the complete absence of up-regulated TFs from several large families (e.g. AP2, TALE, FAR1, G2-like, C3H and MYB-related) at 3-HAI, which uncovered a clearly biased activation of particular families at this stage: 69% (56/81) of the TFs up-regulated at 3-HAI belong to only 12.7% of the TF families that have at least one member transcribed during germination (ERF: 13, LOB: 9, HD-ZIP: 9, bZIP: 7, bHLH: 7, GRAS: 6, Dof: 5).

In addition to TFs, chromatin proteins are also extremely relevant in regulating gene expression on a large scale. We found enrichment for the term GO:0000785 (i.e. chromatin) (Table S3) among genes up-regulated at 24-HAI; this group comprises 41 genes, out of which 36 encode histone subunits and 5 encode high mobility group A genes (HMGA), which are key non-histone proteins that contain AT-hook motifs<sup>39,40</sup>. The up-regulation of these genes late in germination may be important for the chromatin structure changes underlying DNA duplication and cell division that take place during early seedling growth.

**Transcriptional activation of primary metabolism genes during germination.** Although most of the stored reserves are mobilized after seed germination, catabolic pathways must be activated early upon imbibition to sustain axis growth. However, the relative contribution and transcriptional regulation of each pathway to the metabolism of the embryonic axis are not fully understood. Sucrose and raffinose family oligosaccharides (RFOs) are the main soluble carbohydrates available in soybean seeds<sup>41</sup> and are mainly accumulated during desiccation<sup>42</sup>. RFO concentrations sharply decline during germination, concomitantly with a loss of desiccation tolerance and an increase in the levels of reducing sugars<sup>43</sup>. In terms of energy production, invertases are the major drivers of sucrose usage. We found 10 expressed invertases, out of which 6 were induced at 24-HAI and 4 showed high expression levels at this stage (Fig. 2; Table S4). These results support an increasing demand for sucrose-derived monosaccharides later in germination and probably during early seedling growth.

Sucrose can also be cleaved in a reversible reaction catalysed by sucrose synthase (SuSy; Fig. 2). While invertases are associated with high energy demand conditions (e.g. cell expansion), SuSy activity has been linked with the biosynthesis of structural or reserve polysaccharides<sup>44</sup>. Although reversible, SuSy activity favours sucrose cleavage in most heterotrophic tissues, providing sugar-nucleotide precursors for structural and storage polysaccharide biosynthesis<sup>44</sup>. Further, SuSy can associate with membranes<sup>44</sup> and channel UDP-glucose to cellulose synthase complexes<sup>45</sup>, which interact with microtubules and guide cellulose deposition and cellular elongation<sup>46</sup>. Six and 19 SuSy and cellulose synthase genes were highly expressed at 24-HAI, respectively (Fig. 2; Table S4). As discussed above, 14 of these 19 cellulose synthase genes were up-regulated at 24-HAI (Table S3). The complex transcriptional patterns of invertase and SuSy genes probably constitute an intricate system of sucrose partitioning between ATP production from reducing sugars and cellulose biosynthesis from UDP-glucose, which are two essential processes underlying cellular expansion during germination. Finally, in spite of the intense sucrose degradation supported by our results, it has been long shown that sucrose concentrations do not decrease during germination<sup>41,43</sup>, suggesting a renewal of the embryonic axes' sucrose pools, probably from the degradation of other reserves (e.g. oil).

We also found an early activation of glycolysis and respiratory pathways, which could be partially fed by sucrose degradation. Among these genes are: 12 hexokinases; 12 phosphofructokinases (PFK); 13 glyceraldehyde 3-phosphate dehydrogenases (GAPDH); triosephosphate isomerases (TIM) and pyruvate kinases (PKs) (Fig. 2; Table S4). The *A. thaliana* *At2g21170* encodes a plastid TIM that is required for heterotrophic to autotrophic growth in post-germinative stages<sup>47</sup>. We found two highly expressed *At2g21170* orthologs during soybean germination (Fig. 2), suggesting that this gene could also be critical in the transition to autotrophic growth in soybean. Glucose can also be consumed by the pentose phosphate pathway, which has a committed step catalysed



**Figure 2. Major metabolic pathways and enzymes discussed in this study.** Transcriptional levels of key genes are represented in heatmaps. Genes are numbered as following: 1) invertase; 2) sucrose synthase (SuSy); 3) cellulose synthase; 4) hexokinase (HK); 5) glucose-6-phosphate dehydrogenase (G6PDH); 6) phosphofructokinase (PFK); 7) Fructose 1,6-bisphosphatase (FBPase); 8) diphosphate-fructose-6-phosphate 1-phosphotransferase (PFP); 9) pyruvate kinase (PK); 10) pyruvate decarboxylase; 11) alcohol dehydrogenase (ADH); 12) lactate dehydrogenase (LDH); 13) pyruvate dehydrogenase subunit E2 (PDH-E2); 14) citrate synthase (CSY); 15) isocitrate dehydrogenase (IDH); 16) isocitrate lyase (ICL); 17) malate synthase (MSY); 18) malate dehydrogenase (MDH); 19) phosphoenolpyruvate carboxykinase (PEPCK); 20) triacylglycerol (TAG) lipase; 21) glycerol kinase; 22) glycerol-3-phosphate dehydrogenase (GPDHC1); 23) COMATOSE (CTS); 24) glutamate decarboxylase (GAD); 25) gamma-amino-N-butyrate transaminase (GABA-T). The 5 columns in each heatmap represent the germination time points (dry; 3-HAI; 6-HAI; 12-HAI; 24-HAI). Other abbreviations: G3P: glucose-3-phosphate; DHAP: dihydroxyacetone phosphate; PEP: phosphoenolpyruvate; SSA: succinic semialdehyde.

by glucose-6-phosphate dehydrogenase (G6PDH) (Fig. 2). Eight G6PDH genes were transcribed during germination (Fig. 2), indicating that glucose is partially channelled to the pentose phosphate pathway during soybean germination. Further, we found moderate transcription levels of 3 fructose-1,6-bisphosphatases (FBPase), which catalyse the reverse reaction of PFK, converting fructose 1,6-bisphosphate to fructose-6-phosphate in the Calvin Cycle and gluconeogenesis (Fig. 2; Table S4). This interconversion is tightly regulated by fructose 2,6-bisphosphate (Fru2,6P<sub>2</sub>), which activates PFK and represses FBPase in non-plant organisms<sup>48</sup>. In plants, however, Fru2,6P<sub>2</sub> activates a pyrophosphate dependent PFK (PFP), whereas the ATP-dependent PFK is largely insensitive to Fru2,6P<sub>2</sub><sup>48</sup>. Seven PFP genes are expressed during germination, suggesting a complex interplay between PFKs and PFPs (Fig. 2; Table S4). Overall, our results support an intense carbohydrate catabolism via glycolysis/TCA cycle and pentose phosphate pathway. Nevertheless, the expression of FBPase and two critical sucrose biosynthesis enzymes [i.e. sucrose-phosphate synthase (e.g. *Glyma13g23060*) and sucrose-phosphate phosphatase (e.g. *Glyma20g16990*)] should not be neglected, as fluctuations in Fru2,6P<sub>2</sub> concentrations may trigger sucrose

biosynthesis in particular conditions/sites. The expression of sucrose biosynthesis enzymes may contribute to the regeneration of the sucrose pools discussed above.

Acetyl-CoA is central for energy metabolism, particularly in: 1) mitochondria, where it is oxidized via the TCA cycle and; 2) glyoxysomes, where it is an intermediate in the conversion of fatty acids to carbohydrates. Acetyl-CoA can be produced through a transacylation reaction catalysed by the pyruvate dehydrogenase E2 subunit of the pyruvate dehydrogenase complex (PDC). We found 5 PDC-E2 genes for which transcription increases as germination proceeds, suggesting that aerobic metabolism resumes very early in germination (Fig. 2; Table S4). Further, PDH-E1s (19) and PDH-E3s (7) PDC subunits were also highly transcribed, as well as key TCA enzymes, such as citrate synthase (CSY; 13 genes), isocitrate dehydrogenase (IDH; 6 genes) and malate dehydrogenase (MDH; 8 genes) (Fig. 2; Table S4). Importantly, most of these TCA genes reach their maximum transcriptional levels at 24-HAI, which could explain the increased activity of respiratory pathways after mitochondrial repair and biogenesis, which take place early in germination<sup>10,49</sup>.

Despite the rapid water uptake and activation of genes involved in respiratory pathways, oxygen levels are often sub-optimal inside the seeds. Together with the presence of immature or damaged mitochondria in imbibed seeds, fermentation has been suggested to play central roles in the onset of germination<sup>49</sup>. In support of this hypothesis, we found highly expressed pyruvate decarboxylase, alcohol dehydrogenase (ADH) and lactate dehydrogenase (LDH) genes during germination (Fig. 2; Table S4). Strikingly, all these genes reach maximum transcriptional levels at 6-HAI, implying that fermentation is indeed more relevant before mitochondria are fully repaired or synthesized *de novo*.

Triacylglycerols stored in oil bodies are typically located close to glyoxysomes, organelles specialized in converting fatty acids derived from TAGs into soluble-sugar precursors via the glyoxylate cycle<sup>50</sup>. Although the glyoxylate and TCA cycles share several enzymes, the former bypasses the decarboxylation steps of the latter by the activity of isocitrate lyase (ICL) and malate synthase (MSY). Malate is then oxidized to oxaloacetate by malate dehydrogenase (MDH), which is subsequently converted to phosphoenolpyruvate (PEP) by phosphoenolpyruvate carboxykinase (PEPCK)<sup>50</sup> (Fig. 2; Table S4). Interestingly, we found moderate expression levels for two ICL genes and high expression levels for two MSY genes, particularly the glyoxysomal MSY (*Glyma17g13730*) (Fig. 2; Table S4). Three MDH and 3 PEPCK genes reach their maximum expression levels at 24-HAI (Fig. 2; Table S4). Further, at least 6 TAG (GDSL-like) lipase and 4 COMATOSE genes<sup>51</sup> were also expressed during germination, supporting the transport of fatty acyl-CoA into peroxisomes/glyoxysomes (Fig. 2; Table S4). The mobilization of triacylglycerols also generates significant amounts of glycerol, which can enter glycolysis as dihydroxyacetone phosphate (DHAP) via a glycerol shunt pathway involving glycerol kinase (GLI1) and glycerol-3-phosphate dehydrogenase (G3PDc)<sup>10</sup>. Accordingly, we found a GLI1 and 5 GPDHc genes transcribed during germination (Fig. 2; Table S4). It is widely accepted that the mobilization of oil reserves is more critical after germination<sup>50,52</sup>, although several studies demonstrated that peroxisomal  $\beta$ -oxidation and glyoxylate cycle determine germination potential under several conditions<sup>53,54</sup>. Further, the early activation of genes involved in lipid mobilization reported here is in accordance with a previous report that storage oil bodies localized at embryonic cells are locally catabolized during germination<sup>49</sup>.

Seed storage proteins are critical to provide amino acids for protein synthesis and energy production, particularly glutamate and aspartate, which are the most abundant amino acids in soybean seeds (FAO; <http://www.fao.org/docrep/t0532e/t0532e02.htm>). They are substrates of alanine and aspartate aminotransferases (AlaAT and AspAT), enzymes that are activated during imbibition and likely to play important roles in respiratory pathways<sup>10</sup>. Several lines of evidence support the importance of Asp accumulation in germination, such as the negative influence of ABA on AspAT protein levels and fast increase in AspAT transcription after imbibition<sup>10,49</sup>. Our results support the high activity of these enzymes, as 7 AspATs and 4 AlaATs are expressed throughout germination. The non-protein amino acid  $\gamma$ -aminobutyrate (GABA) also has important energetic roles during seed germination<sup>10,49</sup>, as demonstrated in *Arabidopsis*<sup>52</sup> and chestnut<sup>55</sup>. Further, GABA accumulation in germinating soybean seeds is correlated with glutamate decarboxylase (GAD) activity<sup>56</sup>. GAD synthesizes GABA from glutamate and is the first enzyme in the GABA shunt pathway. GABA is then metabolized by GABA transaminases (GABA-T), generating succinic semialdehyde that enters the TCA as succinate (Fig. 2; Table S4). We found expressed GAD and GABA-T genes, supporting the activity of a GABA shunt during soybean germination (Fig. 2; Table S4).

It has been shown for several species (e.g. barley) that amino acids used for *de novo* protein synthesis come from the degradation of storage proteins<sup>49</sup>. However, we found strong transcriptional evidence supporting the activation of some amino acid biosynthesis pathways during germination. A crucial step in the biosynthesis of aromatic amino acids (i.e. phenylalanine, tyrosine and tryptophan) is the production of chorismate via the shikimate pathway, which starts with a rate-limiting reaction catalysed by 3-deoxy-7-phosphoheptulonate (DAHP) synthase: the conversion of PEP and erythrose 4-phosphate (from the pentose phosphate pathway) into 3-deoxy-D-arabino-hept-2-ulosonate 7-phosphate<sup>57</sup>. DAHP synthase links the carbohydrate metabolism to the shikimate pathway, which will ultimately produce chorismate, a central metabolite in the biosynthesis of aromatic amino acids, salicylate and B9 vitamin (folate)<sup>57</sup>. We found high transcriptional levels of a DAHP synthase throughout germination (*Glyma15g06020*, FPKM > 37), supporting that some carbon from the pentose phosphate pathway feed into the chorismate pathway. Other shikimate pathway genes were also highly expressed, such as shikimate kinase (e.g. *Glyma04g39700*) and chorismate synthase (*Glyma20g31980*). Since chorismate is a substrate for the synthesis of different metabolites, we investigated its possible routes in more detail. Firstly, we found 7 expressed chorismate mutase genes (e.g. *Glyma01g06660*), which catalyse the synthesis of prephenate, a substrate for Tyr and Phe biosynthesis. We have also found expression of arogenate dehydrogenases (e.g. *Glyma11g35760*) and arogenate dehydratases (e.g. *Glyma11g15750* and *Glyma17g05290*), which are required to produce Tyr and Phe, respectively. Further, we found two expressed aminodeoxychorismate synthases (i.e. *Glyma10g35580* and *Glyma20g31966*) and one 4-amino-4-deoxychorismate lyase (*Glyma12g35520*), indicating that folate biosynthesis from chorismate is also active during germination. Conversely, no isochorismate synthase

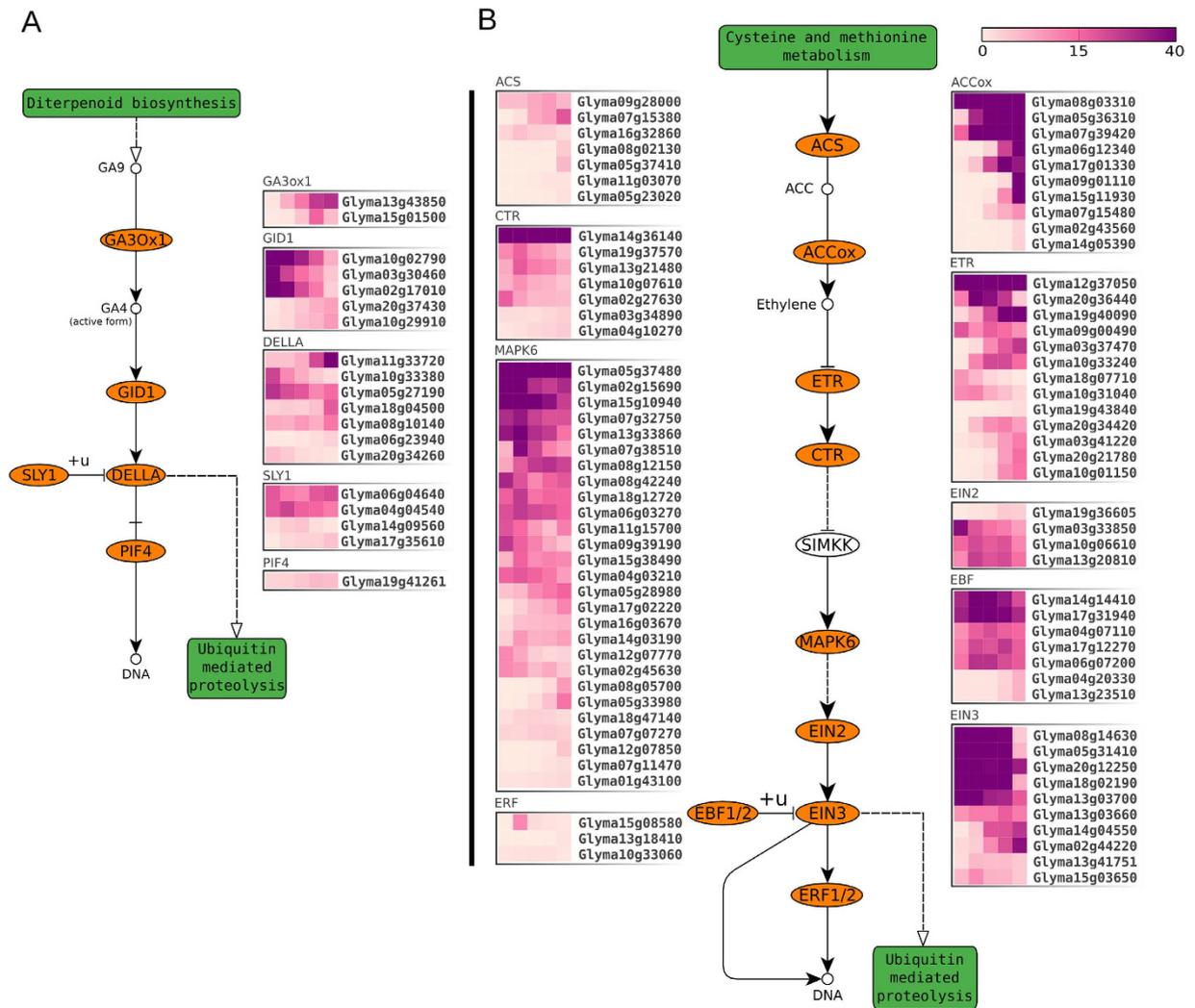
gene is expressed, supporting the absence of salicylate biosynthesis from chorismate during germination. Although nuclear-encoded, the shikimate pathway and other biosynthetic processes take place at the plastids, which are largely non-functional in dry seeds. However, a recent study demonstrated that plastid biogenesis starts early in phase I of soybean germination<sup>58</sup>. Therefore, we hypothesize that newly generated plastids get promptly equipped with the shikimate and other biosynthetic pathways.

**Hormone biosynthesis and signalling during germination.** The embryonic axis gives rise to the organ primordia and has primary regulatory roles in plant growth. In this section we explore the expression of genes involved in the biosynthesis and signalling of key hormones during germination. Seed desiccation and maintenance of dormancy are positively regulated by ABA<sup>8,10</sup>. ABA Insensitive3/Viviparous1 (ABI3/VP1) is a TF that mediates ABA responses by promoting storage reserve deposition, desiccation tolerance and seed dormancy<sup>59</sup>. We found two highly expressed ABI3 genes during germination (*Glyma08g47240* and *Glyma18g38490*, FPKM > 62; Table S1). In *Arabidopsis* and tomato, ABI3 expression is repressed soon after germination by an endospermic GA or GA-derived signal<sup>60</sup>. This also seems to be the case in soybean, as *Glyma08g47240* and *Glyma18g38490* remain highly expressed by 12-HAI, being 9 to 10-fold down-regulated by 24-HAI. ABI3 degradation is promoted by the AIP2 E3 ligase<sup>61</sup>, for which we found two expressed homologs (*Glyma17g33631* and *Glyma14g12380*). Interestingly, ABI5, a TF that acts downstream of ABI3<sup>62</sup>, is also highly expressed early during germination and repressed by 24-HAI. Similarly to ABI3, ABI5 is also down-regulated by proteolysis<sup>63</sup>. These results indicate that ubiquitin-mediated proteolysis of critical TFs involved in ABA response precedes their transcriptional repression, probably allowing a fine-tuned regulation of ABA effects.

ABA catabolism plays important roles in regulating germination potential, particularly by ABA 8'-hydroxylation catalysed by the CYP707A1-4 family<sup>59</sup>. CYP707A1 reduces ABA concentrations during mid-maturation, whereas CYP707A2 is more related to late-maturation and imbibition stages<sup>64</sup>. Accordingly, we found a CYP707A2 gene that is highly expressed in dry seeds and at 3-HAI (*Glyma09g41960*; FPKM > 91), being strongly down-regulated at 24-HAI (FPKM = 2.58). Other CYP707A genes display different expression patterns. Conversely, the CYP707A1 gene *Glyma16g20490* is transcribed below the detection threshold until 6-HAI, being up-regulated at 12 and 24-HAI. Given that *Glyma16g20490* is not expressed in dry seeds, it is likely to be transcriptionally inactive in late-maturation. Therefore, our results reveal the opposed and cyclic transcription of CYP707A1 and CYP707A2 genes during soybean maturation and germination, which could constitute a system to balance active ABA levels. Surprisingly, several other highly expressed ABA-signalling genes were also expressed during soybean germination (e.g. PYL genes: *Glyma16g02910*; HAI3 ABA-induced protein phosphatases: *Glyma01g43460*). Therefore, we hypothesize that post-translational regulation and ABA catabolism are the major mechanisms modulating ABA effects before germination-promoting transcriptional programs are fully installed.

GA is a key hormone that promotes seed germination, largely by antagonizing ABA effects<sup>8</sup>. GA binding to the gibberellin-insensitive dwarf1 (GID1) receptor promotes the formation of a DELLA-binding surface<sup>65,66</sup>. The GA-GID1-DELLA complex is then recognized by SLEEPY1 (SLY1), which is the F-box component of the SCF<sup>SLY1</sup> E3 ubiquitin ligase complex. Interaction with SCF<sup>SLY1</sup> results in DELLA ubiquitination and degradation, releasing GA responses<sup>67</sup>. We found 2 and 5 highly expressed SLY1 and GID1 genes during germination, respectively. Among GID1 genes, 3 are highly transcribed in dry seeds (Fig. 3). Interestingly, these 3 genes belong to the GID1B subfamily, which is induced during after-ripening in the absence of GA biosynthesis<sup>68</sup>. GID1B proteins have higher affinity for GA than GID1A-C<sup>69</sup> and their overexpression increase GA sensitivity and mimic dormancy-breaking treatments more efficiently than that of GID1A-C<sup>68</sup>. Further, GID1A-C and GID1B group in phylogenetically distinct clades and have different expression patterns during germination in Brassicaceae<sup>70</sup>. Our results are coherent with the functional divergence between GID1A-C and GID1B genes, as the two soybean GID1Cs increase in expression as germination proceeds, as opposed to GID1Bs, which are up to 16-fold down-regulated between 0 and 24-HAI (Fig. 3). Since the high expression of GID1B is not able to compensate the absence of GID1A-C<sup>70</sup>, we hypothesize that the increased expression of GID1C genes is critical to mediate the canonical GA response during soybean seed germination. The high abundance of GID1B transcripts in dry seeds, together with their higher affinity for GA, may be part of a mechanism to perceive very low GA levels promptly upon imbibition, as previously hypothesized<sup>68</sup>. Two SLY1 genes are highly transcribed in all time points (Fig. 3), supporting the capacity of the embryonic axes to sense a wide range of GA levels throughout germination. Finally, we found 7 DELLA genes with distinct expression patterns (Fig. 3). The differential regulation of DELLA and GID1 genes suggest a complex regulatory system, especially because DELLA induces the expression of GID1 and GA biosynthesis genes. Accordingly, we found two highly expressed GA3ox1 genes (Fig. 3), which supports the embryonic axis a major GA biosynthesis site during germination<sup>71</sup>. Since GID1 promotes DELLA proteolysis in the presence of GA, DELLA has been proposed as a central player in GA homeostasis via feedback regulation<sup>72</sup>.

Although not as studied as ABA and GA, ethylene and BR have also been shown to promote seed germination<sup>27,73</sup>. As discussed above, important components of the BR pathway are up-regulated during germination (e.g. BRI1). On the other hand, ethylene production starts very early upon imbibition and promotes dormancy release and germination by regulating ABA metabolism and signalling<sup>74,75</sup>. The main intermediate of ethylene production is S-adenosyl-methionine (S-AdoMet), which is synthesized from methionine by S-AdoMet synthetase and converted to 1-aminocyclopropane 1-carboxylic acid (ACC) by ACC synthase (ACS). Ethylene synthesis is then catalysed by ACO, which is critical for germination<sup>76</sup> and presents a high correlation between enzymatic activity and transcript abundance<sup>75</sup>. Further, we found high transcriptional levels of several of these three key ethylene biosynthesis genes throughout germination (Fig. 3). Ethylene can bind to 5 types of transmembrane receptors located at the endoplasmic reticulum: ethylene insensitive 4 (EIN4), ethylene resistant 1 (ETR1), ETR2, ethylene response sensor 1 (ERS1) and ERS2<sup>75</sup>. Genes from all these families, except ERS2, were transcribed in high levels during germination. Further, these results reveal some interesting trends: two ETR1 genes (i.e. *Glyma09g00490*, *Glyma12g37050*) are highly expressed in all stages, whereas ERS1 and ETR2 genes are generally biased towards



**Figure 3.** Last biosynthesis steps and core signalling pathways of the germination promoting hormones gibberellin (A) and ethylene (B). Schemes were retrieved from KEGG (ko04075). Shapes and arrows follow the KEGG representation standards ([http://www.genome.jp/kegg/document/help\\_pathway.html](http://www.genome.jp/kegg/document/help_pathway.html)), except for colour codes. Ellipses represent gene/family. Green rectangles represent connected pathways. Heatmaps represent transcriptional levels (in FPKM) of expressed genes. The 5 columns in each heatmap represent the germination time points (dry; 3-HAI; 6-HAI; 12-HAI; 24-HAI).

the end of germination (Fig. 3) and EIN4 genes display more variable transcriptional profiles (Fig. 3). Ethylene binding to these receptors inactivates CTR1 (constitutive triple response), resulting in the activation of the kinase cascade that controls EIN2 and its nuclear effectors, such as EIN3, EILs and ERFs (ethylene response factors), which will regulate the transcription of ethylene response genes (Fig. 3)<sup>75</sup>. Finally, our results support the feedback regulation in ethylene signalling during soybean germination, as some negative regulators of ethylene response were also highly expressed, such as CTR1 and the F-box EBF1<sup>77</sup> (Fig. 3).

In the present work we described the sequencing and analysis of the embryonic axis transcriptome during soybean germination. We found a remarkable enrichment of TFs among the genes up-regulated in the onset of germination, especially from the LOB domain family. This suggests that a large set of early TFs play critical roles in triggering the initial transcriptional programs required for germination. We found strong evidence for the transcriptional regulation of several primary metabolic pathways involved in reserve mobilization and ATP production, such as the early activation of respiratory pathways and local degradation of oil reserves to feed the glyoxylate pathway, generally believed to be critical only after germination. We also found evidence for aromatic amino acid biosynthesis at the embryonic axes during germination. Further, several genes involved in the biosynthesis and signalling of the germination-promoting hormones GA, ethylene and BR were up-regulated in distinct stages during germination. Genes involved in ABA degradation and negative regulation were also activated, supporting the notion that the GA:ABA ratio and interplay are critical for germination<sup>9</sup>. Further, we found several cases of strong transcriptional divergence between paralogs in these hormonal regulatory pathways, suggesting that gene duplications play important roles in the regulation of germination. Our results also underscore the importance of post-translational regulatory mechanisms via protein phosphorylation and ubiquitination, which

may have important contributions from recent duplications and diversification of F-box genes in the soybean lineage<sup>78</sup>. Taken together, we believe the processes described here shed light on important aspects of the soybean germination and may fuel future biotechnological applications.

## Methods

**Plant material, seed germination and RNA purification.** Soybean seeds (cultivar BRS 284) were sown in Petri dishes (14 cm of diameter) containing 2 g of cotton soaked with sterile distilled water (30 mL). Petri dishes were placed into a climate-controlled growth chamber at 28 °C, 12/12 photo-period and 60% relative humidity. In addition to dry seeds, imbibed seeds were harvested at 3, 6, 12 and 24-HAI. Embryonic axes were carefully separated from cotyledons for RNA extraction and weighing. For RNA extraction, collected embryonic axes were immediately placed in RNAlater (Qiagen) on ice until RNA extraction with the RNeasy Plant Mini Kit (Qiagen). Two biological replicates (20 seeds each) of each condition were used. RNA quality was assessed using 1% agarose gel, Picodrop Microliter spectrophotometer and Agilent Bioanalyzer 2100.

**High-Throughput RNA sequencing and analysis.** RNA-Seq libraries were prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina, cat. RS-122-2001) and submitted to paired-end sequencing (2 × 100 bp) using a HiSeq 2000 instrument at LaCTAD (UNICAMP, Campinas, Brazil). Sequencing quality was assessed with *fastqc* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The first 10 bp of each read were trimmed using Trimmomatic-0.30 (option “headcrop = 10”)<sup>79</sup>. The *G. max* reference genome (version 1.1)<sup>3</sup> was downloaded from Phytozome<sup>80</sup>. Trimmed RNA-Seq reads were mapped to the reference genome using Bowtie 2.1.0<sup>81</sup> and TopHat 2.0.9<sup>82</sup>. TopHat was executed using the “-b2-very-sensitive” parameters (-D 20 -R 3 -N 0 -L 20 -i S,1,0.50).

Transcriptional levels were estimated with Cufflinks (version 2.1.1; default parameters)<sup>83</sup> and normalized by fragments per kilobase of transcript per million mapped reads (FPKM). Genes with FPKM equal or greater than 1 were considered “expressed”. Cuffdiff<sup>83</sup> was used to find: a) up-regulated genes between sequential time points (dry vs. 3-HAI; 3-HAI vs. 6-HAI; 6-HAI vs. 12-HAI; 12-HAI vs. 24-HAI) and; (b) genes preferentially expressed at a particular time point. Up-regulated genes were filtered with the Cumberbund<sup>84</sup> *getSig()* method ( $\alpha = 0.1$ ,  $level = 'genes'$ ); these parameters resulted in differentially expressed genes with at least 2.4 fold-change. Preferentially expressed genes were defined by the *csSpecificity()* function (Cumberbund package), which computes the *Jensen-Shannon distance* (JSD), ranging from 0 (constitutively expressed) to 1 (exclusively expressed). Genes with JSD of 0.7 or greater were defined as “preferentially expressed”. Transcriptional patterns were hierarchically clustered and rendered as heatmaps with Gtools<sup>85</sup>. Additional analyses and parsing steps were performed using in-house perl (<https://www.perl.org/>) and R (<http://www.r-project.org/>) scripts. The transcriptome datasets generated in this study were submitted to GEO under the accession number GSE83481.

**Analysis of transcription factors (TFs) and Gene Ontology (GO) enrichment.** Soybean TFs and family annotations were retrieved from PlantTFDB<sup>34</sup> and manually inspected. Enrichment of GO terms was performed using agriGO (FDR < 0.01)<sup>86</sup>. Genes expressed in at least one time point were used as the background set. The final lists of enriched GO terms were compared with SEACOMPARE (agriGO website). Redundancy of the enriched GO terms was removed using REVIGO<sup>87</sup>.

## References

- Ainsworth, E. A., Yendrek, C. R., Skoneczka, J. A. & Long, S. P. Accelerating yield potential in soybean: potential targets for biotechnological improvement. *Plant Cell Environ* **35**, 38–52, doi: 10.1111/j.1365-3040.2011.02378.x (2012).
- Lewis, G., Schrire, B., Mackinder, B. & Lock, M. *Legumes of the World* (Royal Botanical Gardens, 2005).
- Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183, doi: 10.1038/nature08670 (2010).
- Schmutz, J. *et al.* A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713, doi: 10.1038/ng.3008 (2014).
- Varshney, R. K. *et al.* Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **31**, 240–246, doi: 10.1038/nbt.2491 (2013).
- Bertioli, D. J. *et al.* The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* **48**, 438–446, doi: 10.1038/ng.3517 (2016).
- Goldberg, R. B., Hoschek, G., Tam, S. H., Ditta, G. S. & Breidenbach, R. W. Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. *Dev. Biol.* **83**, 201–217 (1981).
- Bewley, J. D., Bradford, K., Hilhorst, H. & Nonogaki, H. *Seeds: Physiology of Development, Germination and Dormancy*. 3rd edition edn (Springer, 2013).
- Finch-Savage, W. E. & Leubner-Metzger, G. Seed dormancy and the control of germination. *New Phytol.* **171**, 501–523, doi: 10.1111/j.1469-8137.2006.01787.x (2006).
- Weitbrecht, K., Muller, K. & Leubner-Metzger, G. First off the mark: early seed germination. *J. Exp. Bot.* **62**, 3289–3309, doi: 10.1093/jxb/err030 (2011).
- Nambara, E. & Marion-Poll, A. Abscisic acid biosynthesis and catabolism. *Annu. Rev. Plant Biol.* **56**, 165–185, doi: 10.1146/annurev.arplant.56.032604.144046 (2005).
- Nonogaki, H. Seed dormancy and germination-emerging mechanisms and new hypotheses. *Front Plant Sci* **5**, 233, doi: 10.3389/fpls.2014.00233 (2014).
- Ueguchi-Tanaka, M. *et al.* GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* **437**, 693–698, doi: 10.1038/nature04028 (2005).
- Ni, M., Tepperman, J. M. & Quail, P. H. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* **95**, 657–667 (1998).
- Yang, H., Lu, P., Wang, Y. & Ma, H. The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J.* **65**, 503–516, doi: 10.1111/j.1365-313X.2010.04439.x (2011).
- Kudapa, H. *et al.* Comprehensive transcriptome assembly of Chickpea (*Cicer arietinum* L.) using sanger and next generation sequencing platforms: development and applications. *PLoS One* **9**, e86039, doi: 10.1371/journal.pone.0086039 (2014).

17. Libault, M. *et al.* An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. *Plant J.* **63**, 86–99, doi: 10.1111/j.1365-313X.2010.04222.x (2010).
18. Severin, A. J. *et al.* RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* **10**, 160, doi: 10.1186/1471-2229-10-160 (2010).
19. Song, L. *et al.* Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions. *BMC Genomics* **17**, 57, doi: 10.1186/s12864-016-2378-y (2016).
20. Jones, S. I. & Vodkin, L. O. Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS One* **8**, e59270, doi: 10.1371/journal.pone.0059270 (2013).
21. Wang, L. *et al.* RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol.* **14**, 169, doi: 10.1186/1471-2229-14-169 (2014).
22. Rajjou, L. *et al.* The effect of alpha-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* **134**, 1598–1613, doi: 10.1104/pp.103.036293 (2004).
23. Linkies, A. *et al.* Ethylene interacts with abscisic acid to regulate endosperm rupture during germination: a comparative approach using *Lepidium sativum* and *Arabidopsis thaliana*. *Plant Cell* **21**, 3803–3822, doi: 10.1105/tpc.109.070201 (2009).
24. Daniel, B. *et al.* Oxidation of Monolignols by Members of the Berberine Bridge Enzyme Family Suggests a Role in Plant Cell Wall Metabolism. *J. Biol. Chem.* **290**, 18770–18781, doi: 10.1074/jbc.M115.659631 (2015).
25. Nelson, D. C. *et al.* Karrikins enhance light responses during germination and seedling development in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **107**, 7095–7100, doi: 10.1073/pnas.0911635107 (2010).
26. Wang, J. *et al.* Overexpression of a cytosol-localized rhamnose biosynthesis protein encoded by *Arabidopsis* RHM1 gene increases rhamnose content in cell wall. *Plant Physiol. Biochem.* **47**, 86–93, doi: 10.1016/j.plaphy.2008.10.011 (2009).
27. Steber, C. M. & McCourt, P. A role for brassinosteroids in germination in *Arabidopsis*. *Plant Physiol.* **125**, 763–769 (2001).
28. Wang, M., Sun, S., Wu, C., Han, T. & Wang, Q. Isolation and characterization of the brassinosteroid receptor gene (*GmBRI1*) from *Glycine max*. *Int J Mol Sci* **15**, 3871–3888, doi: 10.3390/ijms15033871 (2014).
29. Guo, H. *et al.* Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **106**, 7648–7653, doi: 10.1073/pnas.0812346106 (2009).
30. Hematy, K. *et al.* A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr. Biol.* **17**, 922–931, doi: 10.1016/j.cub.2007.05.018 (2007).
31. Kim, H. *et al.* Ethylene receptor ETHYLENE RECEPTOR1 domain requirements for ethylene responses in *Arabidopsis* seedlings. *Plant Physiol.* **156**, 417–429, doi: 10.1104/pp.110.170621 (2011).
32. Cao, D., Cheng, H., Wu, W., Soo, H. M. & Peng, J. Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in *Arabidopsis*. *Plant Physiol.* **142**, 509–525, doi: 10.1104/pp.106.082289 (2006).
33. Lim, C. W., Yang, S. H., Shin, K. H., Lee, S. C. & Kim, S. H. The *AtLRK10L1.2*, *Arabidopsis* ortholog of wheat *LRK10*, is involved in ABA-mediated signaling and drought resistance. *Plant Cell Rep.* **34**, 447–455, doi: 10.1007/s00299-014-1724-2 (2015).
34. Jin, J. P., Zhang, H., Kong, L., Gao, G. & Luo, J. C. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* **42**, D1182–D1187, doi: 10.1093/nar/gkt1016 (2014).
35. Shuai, B., Reynaga-Pena, C. G. & Springer, P. S. The lateral organ boundaries gene defines a novel, plant-specific gene family. *Plant Physiol.* **129**, 747–761, doi: 10.1104/pp.010926 (2002).
36. Bell, E. M. *et al.* *Arabidopsis* lateral organ boundaries negatively regulates brassinosteroid accumulation to limit growth in organ boundaries. *Proc Natl Acad Sci USA* **109**, 21146–21151, doi: 10.1073/pnas.1210789109 (2012).
37. Ryu, H. & Hwang, I. Brassinosteroids in plant developmental signaling networks. *Journal of Plant Biology* **56**, 267–273, doi: 10.1007/s12374-013-0298-9 (2013).
38. Lee, H. W., Kim, M. J., Kim, N. Y., Lee, S. H. & Kim, J. LBD18 acts as a transcriptional activator that directly binds to the *EXPANSIN14* promoter in promoting lateral root emergence of *Arabidopsis*. *Plant J.* **73**, 212–224, doi: 10.1111/tpj.12013 (2013).
39. Aravind, L. & Landsman, D. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* **26**, 4413–4421 (1998).
40. Pedersen, T. J., Arwood, L. J., Spiker, S., Guiltinan, M. J. & Thompson, W. F. High mobility group chromosomal proteins bind to AT-rich tracts flanking plant genes. *Plant Mol. Biol.* **16**, 95–104 (1991).
41. Kuo, T. M., Doehlert, D. C. & Crawford, C. G. Sugar Metabolism in Germinating Soybean Seeds - Evidence for the Sorbitol Pathway in Soybean Axes. *Plant Physiol.* **93**, 1514–1520, doi: 10.1104/Pp.93.4.1514 (1990).
42. Obendorf, R. L. *et al.* Accumulation of Soluble Carbohydrates during Seed Development and Maturation of Low-Raffinose, Low-Stachyose Soybean. *Crop Sci.* **49**, 329–341, doi: 10.2135/cropsci2008.06.0370 (2009).
43. Koster, K. L. & Leopold, A. C. Sugars and Desiccation Tolerance in Seeds. *Plant Physiol.* **88**, 829–832, doi: 10.1104/Pp.88.3.829 (1988).
44. Winter, H. & Huber, S. C. Regulation of sucrose metabolism in higher plants: Localization and regulation of activity of key enzymes. *Crit. Rev. Biochem. Mol. Biol.* **35**, 253–289, doi: 10.1080/07352680091139178 (2000).
45. Amor, Y., Haigler, C. H., Johnson, S., Wainscott, M. & Delmer, D. P. A Membrane-Associated Form of Sucrose Synthase and Its Potential Role in Synthesis of Cellulose and Callose in Plants. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9353–9357, doi: 10.1073/pnas.92.20.9353 (1995).
46. Paredes, A. R., Somerville, C. R. & Ehrhardt, D. W. Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* **312**, 1491–1495, doi: 10.1126/science.1126551 (2006).
47. Chen, M. & Thelen, J. J. The Plastid Isoform of Triose Phosphate Isomerase Is Required for the Postgerminative Transition from Heterotrophic to Autotrophic Growth in *Arabidopsis*. *Plant Cell* **22**, 77–90, doi: 10.1105/tpc.109.071837 (2010).
48. Nielsen, T. H., Rung, J. H. & Villadsen, D. Fructose-2,6-bisphosphate: a traffic signal in plant metabolism. *Trends Plant Sci.* **9**, 556–563, doi: 10.1016/j.tplants.2004.09.004 (2004).
49. Rosental, L., Nonogaki, H. & Fait, A. Activation and regulation of primary metabolism during seed germination. *Seed Science Research* **24**, 1–15, doi: 10.1017/S0960258513000391 (2014).
50. Graham, I. A. Seed storage oil mobilization. *Annu. Rev. Plant Biol.* **59**, 115–142, doi: 10.1146/annurev.arplant.59.032607.092938 (2008).
51. Russell, L., Larner, V., Kurup, S., Bougourd, S. & Holdsworth, M. The *Arabidopsis* COMATOSE locus regulates germination potential. *Development* **127**, 3759–3767 (2000).
52. Fait, A. *et al.* *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* **142**, 839–854, doi: 10.1104/pp.106.086694 (2006).
53. Pracharoenwattana, I., Cornah, J. E. & Smith, S. M. *Arabidopsis* peroxisomal citrate synthase is required for fatty acid respiration and seed germination. *Plant Cell* **17**, 2037–2048, doi: 10.1105/tpc.105.031856 (2005).
54. Pinfield-Wells, H. *et al.* Sucrose rescues seedling establishment but not germination of *Arabidopsis* mutants disrupted in peroxisomal fatty acid catabolism. *Plant J.* **43**, 861–872, doi: 10.1111/j.1365-313X.2005.02498.x (2005).
55. Desmason, A. M. & Tixier, M. Amino-Acids Content in Germinating-Seeds and Seedlings from *Castanea-Sativa L.* *Plant Physiol.* **81**, 692–695, doi: 10.1104/Pp.81.2.692 (1986).
56. Guo, Y. X., Yang, R. Q., Chen, H., Song, Y. & Gu, Z. X. Accumulation of gamma-aminobutyric acid in germinated soybean (*Glycine max L.*) in relation to glutamate decarboxylase and diamine oxidase activity induced by additives under hypoxia. *Eur. Food Res. Technol.* **234**, 679–687, doi: 10.1007/s00217-012-1678-y (2012).

57. Tzin, V. & Galili, G. The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in *Arabidopsis thaliana*. *Arabidopsis Book* **8**, e0132, doi: 10.1199/tab.0132 (2010).
58. Yu, X. M., Li, A. H. & Li, W. Q. How membranes organize during seed germination: three patterns of dynamic lipid remodelling define chilling resistance and affect plastid biogenesis. *Plant Cell and Environment* **38**, 1391–1403, doi: 10.1111/pce.12494 (2015).
59. Holdsworth, M. J., Bentsink, L. & Soppe, W. J. J. Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytol.* **179**, 33–54, doi: 10.1111/j.1469-8137.2008.02437.x (2008).
60. Bassel, G. W., Mullen, R. T. & Bewley, J. D. ABI3 expression ceases following, but not during, germination of tomato and *Arabidopsis* seeds. *J. Exp. Bot.* **57**, 1291–1297, doi: 10.1093/jxb/erj101 (2006).
61. Zhang, X., Garretton, V. & Chua, N. H. The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes Dev.* **19**, 1532–1543, doi: 10.1101/gad.1318705 (2005).
62. Lopez-Molina, L., Mongrand, S., McLachlin, D. T., Chait, B. T. & Chua, N. H. ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant J.* **32**, 317–328 (2002).
63. Garcia, M. E., Lynch, T., Peeters, J., Snowden, C. & Finkelstein, R. A small plant-specific protein family of ABI five binding proteins (AFPs) regulates stress response in germinating *Arabidopsis* seeds and seedlings. *Plant Mol. Biol.* **67**, 643–658, doi: 10.1007/s11103-008-9344-2 (2008).
64. Okamoto, M. *et al.* CYP707A1 and CYP707A2, which encode abscisic acid 8'-hydroxylases, are indispensable for proper control of seed dormancy and germination in *Arabidopsis*. *Plant Physiol.* **141**, 97–107, doi: 10.1104/pp.106.079475 (2006).
65. Shimada, A. *et al.* Structural basis for gibberellin recognition by its receptor *GID1*. *Nature* **456**, 520–U544, doi: 10.1038/nature07546 (2008).
66. Murase, K., Hirano, Y., Sun, T. P. & Hakoshima, T. Gibberellin-induced DELLA recognition by the gibberellin receptor *GID1*. *Nature* **456**, 459–U415, doi: 10.1038/nature07519 (2008).
67. Tyler, L. *et al.* DELLA proteins and gibberellin-regulated seed germination and floral development in *Arabidopsis*. *Plant Physiol.* **135**, 1008–1019, doi: 10.1104/pp.104.039578 (2004).
68. Hauvermale, A. L., Tuttle, K. M., Takebayashi, Y., Seo, M. & Steber, C. M. Loss of *Arabidopsis thaliana* Seed Dormancy is Associated with Increased Accumulation of the *GID1* GA Hormone Receptors. *Plant and Cell Physiology* **56**, 1773–1785, doi: 10.1093/pcp/pcv084 (2015).
69. Nakajima, M. *et al.* Identification and characterization of *Arabidopsis* gibberellin receptors. *Plant J.* **46**, 880–889, doi: 10.1111/j.1365-313X.2006.02748.x (2006).
70. Voegelé, A., Linkies, A., Müller, K. & Leubner-Metzger, G. Members of the gibberellin receptor gene family *GID1* (*GIBBERELLIN INSENSITIVE DWARF1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination. *J. Exp. Bot.* **62**, 5131–5147, doi: 10.1093/jxb/err214 (2011).
71. Yamaguchi, S., Kamiya, Y. & Sun, T. P. Distinct cell-specific expression patterns of early and late gibberellin biosynthetic genes during *Arabidopsis* seed germination. *Plant J.* **28**, 443–453, doi: 10.1046/j.1365-313X.2001.01168.x (2001).
72. Sun, T. P. Gibberellin-*GID1*-DELLA: A Pivotal Regulatory Module for Plant Growth and Development. *Plant Physiol.* **154**, 567–570, doi: 10.1104/pp.110.161554 (2010).
73. Gianinetti, A., Laarhoven, L. J. J., Persijn, S. T., Harren, F. J. M. & Petruzzelli, L. Ethylene production is associated with germination but not seed dormancy in red rice. *Ann. Bot.* **99**, 735–745, doi: 10.1093/aob/mcm008 (2007).
74. Beaudoin, N., Serizet, C., Gosti, F. & Giraudat, J. Interactions between abscisic acid and ethylene signaling cascades. *Plant Cell* **12**, 1103–1115, doi: 10.1105/tpc.12.7.1103 (2000).
75. Corbineau, F., Xia, Q., Bailly, C. & El-Maarouf-Bouteau, H. Ethylene, a key factor in the regulation of seed dormancy. *Frontiers in Plant Science* **5**, doi: 10.3389/Fpls.2014.00539 (2014).
76. Linkies, A. & Leubner-Metzger, G. Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. *Plant Cell Rep.* **31**, 253–270, doi: 10.1007/s00299-011-1180-1 (2012).
77. Potuschak, T. *et al.* EIN3-dependent regulation of plant ethylene hormone signaling by two *Arabidopsis* F box proteins: EBF1 and EBF2. *Cell* **115**, 679–689, doi: 10.1016/S0092-8674(03)00968-1 (2003).
78. Bellieny-Rabelo, D., Oliveira, A. E. A. & Venancio, T. M. Impact of Whole-Genome and Tandem Duplications in the Expansion and Functional Diversification of the F-Box Family in Legumes (Fabaceae). *Plos One* **8**, doi: 10.1371/journal.pone.0055127 (2013).
79. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, doi: 10.1093/bioinformatics/btu170 (2014).
80. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–1186, doi: 10.1093/nar/gkr944 (2012).
81. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359, doi: 10.1038/nmeth.1923 (2012).
82. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36, doi: 10.1186/gb-2013-14-4-r36 (2013).
83. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578, doi: 10.1038/nprot.2012.016 (2012).
84. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version 2.14.0. (2013).
85. Perez-Llamas, C. & Lopez-Bigas, N. Gitoools: Analysis and Visualisation of Genomic Data Using Interactive Heat-Maps. *Plos One* **6**, doi: 10.1371/journal.pone.0019541 (2011).
86. Du, Z., Zhou, X., Ling, Y., Zhang, Z. H. & Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70, doi: 10.1093/nar/gkq310 (2010).
87. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi: 10.1371/journal.pone.0021800 (2011).

## Acknowledgements

This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; E-26/110.236/2011 and E-26/102.259/2013), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We thank the staff of the Life Sciences Core Facility (LaCTAD) from State University of Campinas (UNICAMP) for library preparation and RNA sequencing.

## Author Contributions

Conceived the study: D.B.-R., A.E.A.O. and T.M.V.; Performed the wet lab experiments: E.A.G.d.O., E.d.S.R., E.P.C.; Analyzed the data: D.B.-R. and T.M.V.; Wrote the manuscript: D.B.-R. and T.M.V.; Funding Acquisition: A.E.A.O. and T.M.V.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Bellieny-Rabelo, D. *et al.* Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. *Sci. Rep.* **6**, 36009; doi: 10.1038/srep36009 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016