# Mass-spectrometry-based draft of the Arabidopsis proteome

**Julia Mergner[1], Martin Frejno[1], Markus List[2], Michael Papacek[3], Xia Chen[4], Ajeet Chaudhary[4], Patroklos Samaras[1], Sandra Richter[5], Hiromasa Shikata[6,7], Maxim Messerer[8], Daniel Lang[8], Stefan Altmann[9], Philipp Cyprys[10], Daniel P. Zolg[1], Toby Mathieson[11], Marcus Bantscheff[11], Rashmi R. Hazarika[12,13], Tobias Schmidt[1], Corinna Dawid[14], Andreas Dunkel[14], Thomas Hofmann[14], Stefanie Sprunck[10], Pascal Falter-Braun[9,15], Frank Johannes[12,13], Klaus F. X. Mayer[8,16], Gerd Jürgens[5], Mathias Wilhelm[1], Jan Baumbach[2], Erwin Grill[3], Kay Schneitz[4], Claus Schwechheimer[6] and Bernhard Kuster[1,13,17] ***

[1]*Chair of Proteomics and Bioanalytics, Technical University of Munich (TUM), Freising, Germany*

[2]*Chair of Experimental Bioinformatics, Technical University of Munich (TUM), Freising, Germany*

[3]*Chair of Botany, Technical University of Munich (TUM), Freising, Germany*

[4]*Plant Developmental Biology, Technical University of Munich (TUM), Freising, Germany*

[5]*Center for Plant Molecular Biology, University of Tübingen, Tübingen, Germany*

[6]*Chair of Plant Systems Biology, Technical University of Munich (TUM), Freising, Germany*

[7]*Devision of Plant Environmental Responses, National Institute for Basic Biology, Okazaki, Japan*

[8]*Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health, Munich-Neuherberg, Germany*

[9]*Institute of Network Biology (INET), Helmholtz Center Munich, German Research Center for Environmental Health, Munich-Neuherberg, Germany*

[10]*Cell Biology and Plant Biochemistry, University of Regensburg, Regensburg, Germany*

[11]*Cellzome GmbH, Heidelberg, Germany*

[12]*Population epigenetics and epigenomics, Technical University of Munich (TUM), Freising, Germany*

[13]*Institute of Advanced Study (IAS), Technical University of Munich (TUM), Freising, Germany*

[14]*Chair of Food Chemistry and Molecular Sensory Science, Technical University of Munich (TUM), Freising, Germany*

[15]*Chair of Micobe-Host Interactions, Ludwigs-Maximilians-University (LMU), Munich, Germany*

[16]*Plant Genome Biology, Technical University of Munich (TUM), Freising, Germany*

[17]*Bavarian Biomolecular Mass Spectrometry Center (BayBioMS), TUM, Freising, Germany*

***Abstract***

Plants are indispensable for life on earth and represent organisms of extreme biological diversity with unique molecular capabilities [1]. Here, we present a quantitative atlas of the transcriptomes, proteomes and phosphoproteomes of 30 tissues of the model plant *Arabidopsis*

*thaliana*. It provides initial answers to how many genes exist as proteins (>18,000), where they are expressed, in which approximate quantities (>6 orders of magnitude dynamic range) and to what extent they are phosphorylated (>43,000 sites). We present examples for how the data may be used, for instance, to discover proteins translated from short open reading frames, to uncover sequence motifs involved in protein expression regulation, to identify tissue-specific protein complexes or phosphorylation-mediated signaling events to name a few. Interactive access to this unique resource for the plant community is provided via ProteomicsDB and ATHENA which include powerful bioinformatics tools to explore and characterize Arabidopsis proteins, their modifications and interplay.

## *Main*

The plant model organism *Arabidopsis thaliana* (AT) has revolutionized our understanding of plant biology and influenced many other areas of the life sciences [1]. Knowledge derived from Arabidopsis has also provided mechanistic understanding of important agronomic traits in crop species [2]. The Arabidopsis genome was sequenced 20 years ago and hundreds of natural variants have since been analyzed at the genome and epigenome level [3,4]. In contrast, the Arabidopsis proteome as the main executer of most biological processes is far less comprehensively characterized. To address this gap, we used state-of-the-art mass spectrometry and RNA sequencing (RNA-seq) to provide the first integrated proteomic, phosphoproteomic and transcriptomic atlas of Arabidopsis. Illustrated by selected examples, we show how this rich molecular resource can be used to explore the function of single proteins or entire pathways across multiple omics levels.

## *Multi-omics atlas of Arabidopsis*

We generated an expression atlas covering, on average, 17,603 ± 1,317 transcripts, 14,430 ± 911 proteins and 14,689 ± 2,509 phosphorylation sites (p-sites) per tissue, using a reproducible biochemical and analytical approach (Fig. 1a,b; Extended Data Fig. 1a-c; Supplementary Data 1,2). In total, the protein expression data covers 18,210 of the 27,655 protein-coding genes (66%) annotated in Araport11 [5]. This is a substantial increase compared to the percentage of genes with protein level evidence reported in UniProt (27%) [6] and more than double the number of proteins identified in an earlier tissue proteome analysis [7] (Fig. 1c, Extended Data Fig. 1d-f). In addition, we report tissue-resolved quantitative evidence for a total of 43,903 p-sites making this study the most comprehensive single Arabidopsis phosphoproteome published to date (Fig. 1c). 47% of the expressed proteome was found to be phosphorylated in at least one instance, confirming earlier analyses of individual

phosphoproteomes or information assembled in the PhosPhAT database (Extended Data Fig. 1g,h) [8,9]. The actual figures *in plantae* are likely considerably higher as there is limited overlap between these data sources. This can be readily explained by technical factors (e. g. insufficient sequence coverage) and, more importantly, differences in plant genotypes or stimuli that influence phosphorylation. The above figures are similar to those reported for mammalian systems, underscoring the strong conservation of protein post-translational modifications as a means to create functional diversity from a limited set of genes and proteins [10].

The atlas can be explored using the new web portal ATHENA (Arabidopsis Thaliana ExpressioN Atlas; athena.proteomics.wzw.tum.de) and ProteomicsDB (www.proteomicsdb.org) [11]. Analysis and visualization options include co-expression analysis, the exploration of tissue-specific interaction networks and a tool for predicting tandem mass spectra of peptide sequences, which can be used to validate peptide identifications reported in this study [12] (Fig. 2a,b; Supplementary Table 1). ProteomicsDB further provides spectral libraries for >328,000 unmodified and >43,000 phosphorylated peptides. These can facilitate the development of quantitative mass spectrometry-based protein assays which is particularly useful for research in Arabidopsis, where comparatively few antibodies exist.

***Proteomic annotation of the genome***

Using Araport11, the proteomic data corroborated a substantial number of annotated open reading frame (ORF) borders based on the detection of 2,776 N-terminal and 2,656 C-terminal peptides (Extended Data Fig. 2a). As expected, N-terminal peptides often showed cleavage of the initiator methionine and N-terminal acetylation which was strongly dependent on the amino acid adjacent to the initiator methionine (Extended Data Fig. 2b,c) [13]. The MS data covered, on average, 44% of each protein sequence, enabling the detection of unique peptides for 14,115 protein isoforms. In 297 cases, the data distinguished also two or three splice variants (Extended Data Fig. 2d-f; Supplementary Data 3). A selection of these isoform-specific peptides was validated using synthetic peptides (confirmation rate 80%; Extended Data Fig. 2g). To investigate the potential translation of predicted short open reading frames (sORF), we searched the proteomic data against a published and an in-house compiled sORF collection (ARA-PEP [14]; ATSO) leading to the identification of 51 distinct sORFs that were subsequently confirmed using synthetic peptides (Extended Data Fig. 2g,h; Supplementary Data 3). These results demonstrate the potential of the atlas to refine Arabidopsis gene models and protein sequences.

***Quantitative expression landscapes***

The dynamic range of protein and transcript expression spanned six and four orders of magnitude, respectively (Extended Data Fig. 2i,j). As described before, protein evidence was underrepresented for low abundant transcripts [15], indicating that proteome coverage was not yet exhaustive. Low abundant transcripts were enriched for gene ontology (GO) terms such as cell signaling or gene expression regulation (Supplementary Data 4). Still, the protein data cover ~50% of all annotated transcription factors and ~75% of all transcriptional regulators, kinases and phosphatases, many of which are themselves post-translationally regulated (Extended Data Fig. 2k) [16,17]. Interestingly, protein phosphorylation was detected across the entire dynamic range of protein expression, highlighting the efficiency of the employed phosphopeptide enrichment (Extended Data Fig. 2j).

A breakdown of the molecular data for individual tissues or morphologically similar groups (e.g. leaf, seed or root) showed that only few transcripts or proteins are expressed in a truly specific manner (Extended Data Fig. 3a). Instead, tissue types are characterized by distinct quantitative abundance patterns. For instance, the floral organs express a similar set of genes but at levels that are characteristic for each flower compartment. Here, the MADS domain transcription factors showed the expected differential expression, thus recapitulating established models of floral organ identity [18] (Extended Data Fig. 3b,c).

The strong variation in quantitative gene expression between tissue groups was evident on both protein and transcript level (Extended Data Fig. 3d,e). Still, we found surprisingly little overlap in the rank order or even identity of the most abundant proteins and mRNAs within any given tissue (Extended Data Fig. 3e,f). Extreme expression ranges were detected for e.g. seeds, where the storage protein CRA1 accounted for ~10% of the total protein and the ten most abundant gene products already represented more than 30% of the total amount of protein or transcript, respectively (Extended Data Fig. 3e,g). Similarly, the RuBisCO complex, renowned as the most abundant plant protein on earth, accounted for 4-7% of total leaf protein content [19] (Extended Data Fig. 3g). We note that despite the dominance of some proteins in certain tissues, we show that proteomes of plant tissues can be probed very deeply using current technology, countering the long-standing notion that plant proteomics is particularly difficult [20]. As expected, photosynthetic activity was one of the main factors for expression pattern variations between tissues, which can also be related to plastid localized proteins and their abundance (Extended Data Fig. 3h,i). The most divergent expression pattern on both transcript and protein level, however, was found for pollen (Extended Data Fig. 3h).

The comparison of expression patterns between tissues can be used to ascertain biological functions of individual proteins or to disentangle closely related protein family members. A closer look at the AGCVIII kinase family (23 members in Arabidopsis) [21] revealed high protein expression of members of the D6PK subfamily in embryos as well as of AGC1.5 and AGC1.7 kinases in flowers and pollen (Fig. 2c). In line with these results, follow-up experiments showed promotor activity for *AGC1.5* and *AGC1.7* but not for the closely related *AGC1.6* in pollen, strongly supporting the contribution of the former two in pollen tube growth [22] (Fig. 2d,e). We further uncovered an important role for the D6PKs in embryo development, where a combined knockout synergistically increased the occurrence of aberrant embryo phenotypes in loss-of-function mutants (Fig. 2f-i).

***Regulation of protein amount in tissues***

The amount of protein for a given gene is determined by a regulatory system comprising anabolic and catabolic transcriptional, translational and post-translational processes. This has been extensively studied in unicellular organisms and, more recently, in human [23]. While this broad area of research cannot be covered in detail here, consistent observations were made in the current study. We found positive correlation between transcript and protein levels in most tissues (Pearson's *R* 0.28-0.7, Extended Data Fig. 4a), with the majority of high or low abundant transcripts resulting in high or low abundant proteins, respectively [24] (Fig. 3a). To identify further molecular determinants of protein abundance, we implemented a model selection approach and tested each of the selected features for the potential to explain protein level variations on a tissue-specific or global level (Extended Data Fig. 4b-e; Supplementary Data 5). In addition to well-known predictors like transcript levels and codon usage [25], evolutionary conservation, mRNA sequence motifs and the number of protein interactions were significant predictors of protein levels (Fig. 3b). However, a large proportion (48%) of protein abundance variation still remains unexplained, suggesting that many additional molecular factors with great individual but lower global impact may be at work which will require careful case by case experiments to resolve.

Protein versus transcript abundance plots clearly show that similarly abundant transcripts often lead to proteins with >100-fold differences in abundance. These differences in the protein-to-mRNA ratio (PTR) of genes might be due to differences in translation efficiency of a given transcript, transcript stability and/or the stability of the respective protein (Fig. 3c) [26]. High PTRs were detected for genes involved in photosynthesis or energy metabolism (e.g. petA and rbcL) (Supplementary Data 4). This implies optimized translation and stability of those proteins (and

transcripts) that are needed in large quantities. Conversely, genes with low PTRs were enriched in processes such as hormone-mediated signaling pathways, which, in plants, often involve protein degradation [27]. Members of the auxin-labile AUX/IAA protein family (e. g. IAA8, IAA13) for example show low protein signal despite high transcript levels and are known to be under tight control of auxin-dependent proteasomal degradation [28] (Fig. 3a).

We also observed inter-tissue variation of PTRs, especially for seed and pollen (Fig. 3d). Genes with low PTRs in seeds mainly show differential abundance at the protein level and proteome-wide time-resolved measurements in germinating seeds in response to cycloheximide (translation block) or MG132 (proteasome block) treatment suggest that low PTR proteins are rapidly translated from stored mRNA upon germination. High PTR proteins do not show this behavior implying that they are stored in seed to be readily available for germination (Fig. 3e,f; Extended Data Fig. 4f,g). This is consistent with reports that seeds and pollen accumulate storage molecules including mRNAs and proteins to allow rapid development at the onset of germination and likely explains the observed uncoupling of transcription and translation in these durable plant tissues [29]. Fluctuating PTRs of genes between tissues indicate tissue-specific regulation on either transcript or protein level. Conversely, genes with stable PTRs, are likely under similar regulation in different tissues and the extensive transcriptomic resources available for Arabidopsis may therefore be used to estimate relative protein abundances (Extended Data Fig. 5a,b). Similarly, stable ratios between p-site abundance and the corresponding protein levels, indicates that, under steady-state conditions, the majority of phosphorylation abundance changes between tissues can be attributed to fluctuations of the underlying protein abundance rather than phosphorylation stoichiometry (Extended Data Fig. 5c,d).

***Co-expression of paralogs and complexes***

Based on the 'guilt by association' idea, gene co-expression analysis is widely used to ascribe common functionality between genes and can be performed on transcript or protein level. We observed that co-expressed gene pairs also have high STRING [30] scores, indicative of physical or functional interactions (Extended Data Fig. 6a). To illustrate this point, we focused on co-expression analyses of gene paralogs and interacting proteins. Arabidopsis, like most plants, has undergone genome duplications and retains a considerable number of highly conserved paralogs in the genome [31]. Compared to randomly selected gene pairs, duplicated genes show a clear shift towards positive expression correlations (both transcript and protein). This indicator for redundant gene function is supported by the finding that knock-out mutants of the more abundant paralog are more likely to display a measureable plant phenotype [32] (Extended Data

Fig. 6b-d; Supplementary Data 6). Such comparative expression analyses can be useful to prioritize knock-out combinations for phenotype analyses of functionally redundant genes.

Analyzing proteins that engage in physical interactions recorded in AtPIN [33] for co-expression across tissues suggests that ~26% of AtPIN pairs may be stable rather than transient complex partners (online methods; Extended Data Fig. 6e,f). As expected, protein-based correlations were consistently higher than transcript correlations. Because co-expression analyses tend to generate a large number of interaction candidates, we sought to prioritize these using data from size-exclusion chromatography-mass spectrometry experiments (SEC-MS) as an independent experimental approach [34] (Extended Data Fig. 7a-c; Supplementary Data 7,8). Indeed, we found that co-detection in SEC-MS, in addition to co-expression in the tissue atlas, substantially reduced the list of candidates, illustrated here by the detection of the coatomer complex in the intersection of the SEC-MS and tissue atlas data (Fig. 4a,b). The identification of known protein complexes was consistently improved by using this approach, which combines information from SEC and the tissue atlas datasets (Extended Data Fig. 7d, Supplementary Data 9, see online methods).

Interestingly, in seed tissues, the ζ-subunit of the coatomer complex was almost exclusively provided by the ζ-1 paralog, while this was almost absent in all other tissues (Extended Data Fig. 7e). Subunit abundance comparisons also provided information about complex stoichiometry and similar results were obtained for stable complexes when using the SEC-MS or tissue atlas dataset (Fig. 4c, Extended Data Fig. 7f,g). We, therefore, propose that our tissue atlas resource can provide an initial approximation of the relative subunit frequency of known complexes with unresolved stoichiometry. This could be particularly useful for membrane- or cell wall-associated complexes, which are difficult to study biochemically.

***The Arabidopsis phosphoproteome***

The number of kinases (642 ± 55) and phosphatases (119 ± 6) as well as the ratios of individual families were comparable across most tissues. Pollen and egg cell-like callus, however, stood out implying extensive signaling activity in these tissues (Extended Data Figure 8a-c). In contrast to the wealth of available phosphoproteomic data to date, information on kinase-substrate relationships, critical for understanding signaling cascades or pathway activities, is still sparse. Because these relationships are difficult to discover experimentally, scientists often initially take a computational approach. We used the motif-X algorithm [35] to identify 266 phosphorylation motifs that grouped into ‘proline-directed’, ‘acidic’, ‘basic’ or ‘other’ motif classes [36] (Extended Data Figure 8d-f; Supplementary Table 2, Supplementary Data 10). Together with

information about the co-occurrence of kinases and p-sites in tissues, as well as external information like co-localization or interaction, we anticipate that this dataset can help untangle kinase-substrate and kinase-p-site relationships in the future.

Overall, the number of p-sites per protein showed vast variations. For example, members of the LEA protein family [37] are phosphorylated at almost every serine, threonine or tyrosine residue in their sequences (Fig. 5a; Extended Data Fig. 8g,h). Phosphorylation of these unstructured proteins involved in seed maturation and desiccation may be a mechanism to regulate their conformational state or phase transitions [38]. Conversely, other proteins showed p-site clusters in regulatory domains, including the juxtamembrane domain of receptor-like kinases [39] (Extended Data Figure 8i), implying a role in recruiting interaction partners akin to human receptor tyrosine kinases.

While the mere detection of a phosphorylation event does not directly imply a functional consequence, it provides important starting points [40]. For example, the abscisic acid (ABA) receptor RCAR10, involved in the ABA signaling cascade, was found to be phosphorylated at four sites. We generated phosphomimic mutants (S113D, S32D) that displayed reduced ABA response compared to the wild type receptor (Fig. 5b). While this might be linked to an effect on ABA binding or PP2C interaction for pS113 which is part of the interaction surface for protein phosphatase 2 (PP2Cs) co-receptors, it is more likely due to altered interactions with other regulatory proteins for pS32, which is localized at the receptor periphery [41] (Extended Data Fig. 9a,b). Interestingly, both RCAR10 phosphomimic variants altered the ABA response of RCAR10-PP2C co-receptor combinations, suggesting changes in the interaction profile and phosphorylation-based fine-tuning of ABA signaling during development and stress response (Extended Data Fig. 9c,d).

As a second example for functional consequences of protein phosphorylation, we studied QKY, a protein localized to plasmodesmata and involved in floral and silique morphogenesis [42]. Since one of its p-sites (S262) was detected in all assayed tissues and further p-sites were found between the first and second C2 domains in other MCTP family members, we speculated about a possible effect of pS262 on QKY function (Extended Data Fig. 9e,f). To test this hypothesis, we generated plant lines expressing phosphomimic and phosphomutant *QKY* transgenes in the strong *qky-9* mutant background [43]. The data show that phosphomimic but not phosphomutant protein constructs rescued both flower and silique mutant phenotypes, suggesting that phosphorylation of S262 is required for normal QKY function (Fig. 5c-f, Extended Data Fig. 9g-j).

Taken together, we have generated the most comprehensive, albeit still incomplete, draft of the Arabidopsis (phospho)proteome to date and highlighted some of the many uses that can be envisioned for this resource. Yet, much remains to be accomplished including a more systematic coverage of protein sequence variants or post-translational modifications. The tens of thousands of newly discovered phosphorylation sites that await functional characterization present a particular future challenge. The examples we provide demonstrate that investigating different levels of omics data can lead to new insights into biological processes. We therefore anticipate that this resource along with the provided online computational tools will enable the Arabidopsis research community to perform many further types of systems-level analyses not covered here. We also expect that the study will more broadly impact plant research as the work demonstrates that mass spectrometry-based quantitative protein assays are a veritable way to overcome the disabling lack of antibodies for plant research.

## References


1 Kramer, U. Planting molecular functions in an ecological context with Arabidopsis thaliana. *Elife* **4**, doi:10.7554/eLife.06100 (2015).

2 Peng, J. *et al.* 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256-261, doi:10.1038/22307 (1999).

3 Arabidopsis Genome, I. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796-815, doi:10.1038/35048692 (2000).

4 Kawakatsu, T. *et al.* Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell* **166**, 492-505, doi:10.1016/j.cell.2016.06.044 (2016).

5 Cheng, C. Y. *et al.* Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J* **89**, 789-804, doi:10.1111/tpj.13415 (2017).

6 The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

7 Baerenfaller, K. *et al.* Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. *Science* **320**, 938-941, doi:10.1126/science.1157956 (2008).

8 van Wijk, K. J., Friso, G., Walther, D. & Schulze, W. X. Meta-Analysis of Arabidopsis thaliana Phospho-Proteomics Data Reveals Compartmentalization of Phosphorylation Motifs. *Plant Cell* **26**, 2367-2389, doi:10.1105/tpc.114.125815 (2014).

9 Durek, P. *et al.* PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. *Nucleic Acids Res* **38**, D828-834, doi:10.1093/nar/gkp810 (2010).

10 Sharma, K. *et al.* Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* **8**, 1583-1594, doi:10.1016/j.celrep.2014.07.036 (2014).

11 Schmidt, T. *et al.* ProteomicsDB. *Nucleic Acids Res* **46**, D1271-D1281, doi:10.1093/nar/gkx1029 (2018).

12 Gessulat, S. *et al.* Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* **16**, 509-518, doi:10.1038/s41592-019-0426-7 (2019).

13 Bienvenut, W. V. *et al.* Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol Cell Proteomics* **11**, M111 015131, doi:10.1074/mcp.M111.015131 (2012).

14 Hazarika, R. R. *et al.* ARA-PEPs: a repository of putative sORF-encoded peptides in Arabidopsis thaliana. *BMC Bioinformatics* **18**, 37, doi:10.1186/s12859-016-1458-y (2017).

15 Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587, doi:10.1038/nature13319 (2014).

16 Zheng, Y. *et al.* iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Mol Plant* **9**, 1667-1670, doi:10.1016/j.molp.2016.09.014 (2016).

17 Yang, M. *et al.* A comprehensive analysis of protein phosphatases in rice and Arabidopsis. *Plant Systematics and Evolution* **289**, 111-126, doi:10.1007/s00606-010-0336-8 (2010).

18 Litt, A. & Kramer, E. M. The ABC model and the diversification of floral organ identity. *Semin Cell Dev Biol* **21**, 129-137, doi:10.1016/j.semcdb.2009.11.019 (2010).

19 Bar-On, Y. M. & Milo, R. The global mass and average rate of rubisco. *Proc Natl Acad Sci U S A* **116**, 4738-4743, doi:10.1073/pnas.1816654116 (2019).

20 Gupta, R. *et al.* Time to dig deep into the plant proteome: a hunt for low-abundance proteins. *Front Plant Sci* **6**, 22, doi:10.3389/fpls.2015.00022 (2015).

21 Galvan-Ampudia, C. S. & Offringa, R. Plant evolution: AGC kinases tell the auxin tale. *Trends Plant Sci* **12**, 541-547, doi:10.1016/j.tplants.2007.10.004 (2007).

22 Zhang, Y., He, J. & McCormick, S. Two Arabidopsis AGC kinases are critical for the polarized growth of pollen tubes. *Plant J* **58**, 474-484, doi:10.1111/j.1365-313X.2009.03792.x (2009).

23 Eraslan, B. *et al.* Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol Syst Biol* **15**, e8513, doi:10.15252/msb.20188513 (2019).

24 Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535-550, doi:10.1016/j.cell.2016.03.014 (2016).

25 Hanson, G. & Coller, J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* **19**, 20-30, doi:10.1038/nrm.2017.91 (2018).

26 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).

27 Santner, A. & Estelle, M. The ubiquitin-proteasome system regulates plant hormone signaling. *Plant J* **61**, 1029-1040, doi:10.1111/j.1365-313X.2010.04112.x (2010).

28 Luo, J., Zhou, J. J. & Zhang, J. Z. Aux/IAA Gene Family in Plants: Molecular Structure, Regulation, and Function. *Int J Mol Sci* **19**, doi:10.3390/ijms19010259 (2018).

29 Bai, B. *et al.* Seed stored mRNAs that are specifically associated to monosome are translationally regulated during germination. *Plant Physiol*, doi:10.1104/pp.19.00644 (2019).

30 Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362-D368, doi:10.1093/nar/gkw937 (2017).

31 Wang, Y., Tan, X. & Paterson, A. H. Different patterns of gene structure divergence following gene duplication in Arabidopsis. *BMC Genomics* **14**, 652, doi:10.1186/1471-2164-14-652 (2013).

32 Lloyd, J. & Meinke, D. A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiol* **158**, 1115-1129, doi:10.1104/pp.111.192393 (2012).

33 Brandao, M. M., Dantas, L. L. & Silva-Filho, M. C. AtPIN: Arabidopsis thaliana protein interaction network. *BMC Bioinformatics* **10**, 454, doi:10.1186/1471-2105-10-454 (2009).

34 Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nat Methods* **9**, 907-909, doi:10.1038/nmeth.2131 (2012).

35 Schwartz,D & Gygi, SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnology* **23**1391-1398, doi:10.1038/nbt1146 (2005).

36 Villen, J., Beausoleil, S. A., Gerber, S. A. & Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **104**, 1488-1493, doi:10.1073/pnas.0609836104 (2007).

37 Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F. & Covarrubias, A. A. The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* **148**, 6-24, doi:10.1104/pp.108.120725 (2008).

38 Bah, A. *et al.* Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106-109, doi:10.1038/nature13999 (2015).

39 Mitra, S. K. *et al.* An autophosphorylation site database for leucine-rich repeat receptor-like kinases in Arabidopsis thaliana. *Plant J* **82**, 1042-1060, doi:10.1111/tpj.12863 (2015).

40 Landry, C. R., Levy, E. D. & Michnick, S. W. Weak functional constraints on phosphoproteomes. *Trends Genet* **25**, 193-197, doi:10.1016/j.tig.2009.03.003 (2009).

41 Hauser, F., Li, Z., Waadt, R. & Schroeder, J. I. SnapShot: Abscisic Acid Signaling. *Cell* **171**, 1708-1708 e1700, doi:10.1016/j.cell.2017.11.045 (2017).

42 Vaddepalli, P. *et al.* The C2-domain protein QUIRKY and the receptor-like kinase STRUBBELIG localize to plasmodesmata and mediate tissue morphogenesis in Arabidopsis thaliana. *Development* **141**, 4139-4148, doi:10.1242/dev.113878 (2014).

43 Fulton, L. *et al.* DETORQUEO, QUIRKY, and ZERZAUST represent novel components involved in organ development mediated by the receptor-like kinase STRUBBELIG in Arabidopsis thaliana. *PLoS Genet* **5**, e1000355, doi:10.1371/journal.pgen.1000355 (2009).

373

## Figure legends

**Figure 1 | Tissue map and multi-omics dataset.** a, Schematic representation of tissue samples analysed in this study and coloured according to morphology group: Flower (light grey), seed (dark brown), pollen (yellow), stem (dark green), leaf (light green), root (dark grey), fruit (light brown), callus (magenta), cell culture (blue). b, Number of identifications at the protein, transcript and phosphorylation site (p-site) level for all tissues (n = 1 measurements per tissue). A dashed line indicates the number of core proteins, transcripts or p-sites detected in all tissues. Tissue-enhanced proteins or transcripts are marked in darker colour. P-sites with high confidence amino acid localization (class I,>0.75 localization probability) are shown in pink, ambiguous site localizations in purple. The number of p-sites exclusively detected in one tissue is shown as circles. c, Total number and overlap of identified gene loci in the transcriptome, proteome and phosphoproteome datasets compared to Araport11 (left panel) and total number of identified p-sites and proportion of class I sites (right panel).

**Figure 2 | Data exploration in ATHENA and ProteomicsDB.** a, Schematic representation of data analysis options in the web tool ATHENA (Arabidopsis THaliana ExpressioN Atlas) and ProteomicsDB. b, Mirror plot showing near identical observed and predicted tandem mass spectra (n = 1 acquired spectra) for a peptide from the 'uncertain' protein AT4G13955, thus validating its existence as a protein. SA denotes the normalized spectral contrast angle, a measure for spectrum similarity. c, Upper panel: summed absolute protein expression of AGC1 and AGC3 subfamilies of AGCVIII kinases. Lower panel: relative protein expression profiles of D6PK family members and the closely related kinases AGC1.5, AGC1.6 and AGC1.7. Arrow heads indicate embryo and pollen tissue with high expression of D6PKL3 or AGC1.5 and AGC1.7, respectively. d-e, Representative images of GUS-stained flowers expressing *AGC1.6* (n = 20 independent transgenic lines) and *AGC1.7* (n = 12 independent transgenic lines) promotor-GUS fusions. Arrow-heads indicate anthers containing pollen grains. Scale bars 1 mm. f-i, Representative images of embryo phenotypes (at heart stage or age matched equivalent) and their respective frequency for wild type (WT), *d6pkl3* single, *d6pk012* triple (*d6pk d6pkl1 d6pkl2*) and *d6pk0123* (*d6pk d6pkl1 d6pkl2 d6pkl3*) quadruple knock-out mutants. Scale bars 20 µm.

**Figure 3 | Protein and mRNA expression.** a, Scatter plot of protein versus transcript abundance (median across 30 tissues). Arrows mark examples for genes with high protein to transcript ratio (PTR, rbcL, petA) or low PTR (IAA8, IAA13). Pearson correlation (*R*). b, Relative contribution of molecular features (averaged across tissues) able to predict protein abundance. UTR, untranslated region, Dn/Ds, ratio of non-synonymous to synonymous nucleotide substitutions. c, Schematic representation of the steps where modulation of PTRs may occur. For example, the higher the transcription rate the lower the PTR. d, PTR distribution of genes for selected tissues. Genes with high and low PTRs are defined as being outside two standard deviations of the median PTR distribution. e, Analysis of the proportion of genes with high or low

PTRs that are at least 4-fold differently abundant on transcript and/or protein level in seed compared to all other tissues (depicted as arrows). The percentage of genes below the 4-fold criterion are placed at the origin of the plot. n denotes the number of genes with high or low PTRs in seed as defined in panel d. f, Relative protein expression changes of proteins with high or low PTRs (from panel e) in seeds incubated for 8h, 16h and 24h with 100 µM cycloheximide (translation block), MG132 (proteasome block) or DMSO (control).

**Figure 4 | Protein complex characterization by protein co-expression and SEC-MS.** a, Upper panel: 221 proteins that form a module in the tissue atlas based on their co-expression across all tissues. Lower panel: co-elution of 57 proteins in a size-exclusion chromatography (SEC)-mass spectrometry experiment using flower tissue. b, Overlap of proteins (n=14) between the groups of co-expressed and co-eluting proteins in a. Ten of the proteins in the intersection are subunits of the coatomer complex. c, Tissue-resolved absolute abundance (top panel) and relative proportions of coatomer complex subunits based on the tissue atlas protein expression data (lower panel) showing heterogeneous expression but a fairly homogeneous composition of the complex across tissues. Paralogs of the α, β, β', ε and ζ coatomer subunits were combined for the analysis. Tissue groups are coloured as in Figure 1.

**Figure 5 | Ascribing function to protein phosphorylation.** a, Scatter plot of the proportion of potential phosphorylation acceptor sites in a protein sequence (limited to S, T and Y) versus the proportion of STY residues found to be phosphorylated in this study. While most proteins including RCAR10 and QKY show only a small proportion of phosphorylated residues, the LEA proteins are phosphorylated at almost every possible residue. b, Abscisic acid (ABA) response (average ± SD; n = 3) of ABA treated protoplast cells transformed with wild type or phosphomimic variants of RCAR10 (S32D and S113D) suggesting that S32 and S113 phosphorylation are functionally involved in RCAR10 activity. c-f, Representative images of flowers and siliques of wild type (WT), *qky-9* mutant and *qky-9* expressing phospho-dead $QKY_{S262A}$ or phosphomimic $QKY_{S262E}$ constructs showing that $QKY_{S262E}$ but not $QKY_{S262A}$ rescues the *qky-9* phenotype. Scale bars: flower 1 mm, silique 0.5 cm.

## Methods

### *Sample preparation*

***Inflorescence and seed samples:*** *Arabidopsis thaliana* wild type Columbia-0 (Col-0) plants were grown under continuous white light conditions at 22°C. Samples for flower parts and siliques were harvested from mature plants. Stage 15 flowers [46] were dissected into petal, sepal, stamen and carpel. Fully grown green siliques were separated into septum, valves and green seeds (stage 10). Surface-sterilized mature dry seeds were stored for two days at 4°C and subsequently imbibed for 24 h or kept dry. For pollen collection, plants were grown on soil under a long photoperiod (16 h of light, 21°C, 65% humidity). Mature pollen was bulk-harvested from open flowers at developmental stage 13 [46]. A vacuum cleaner was modified with three subsequent nylon meshes (80 micron, 35 micron, 10 micron mesh) for large scale pollen isolation as described [47]. Pollen was collected in a 1.5 ml reaction tube, snap-frozen in liquid nitrogen and stored at -80°C until further use.

***Cell culture and callus samples:*** Cell culture samples (root cells, Col-0) were grown in medium composed of 4.3 g/l Murashige and Skoog basal medium (MS), 30 g/l sucrose and 0.33 mg/ml $KH_2PO_4$, supplemented with 2.4-D (final concentration 1 mg/l; pH adjusted to 5.8 with KOH) at 22°C under continuous light and harvested either three or ten days after sub-culturing. To generate callus inducing medium (CIM) callus, root explants were harvested from two weeks-old seedlings (Col-0) grown in sterile culture on MS plates. 5 to 10 mm long root segments were cultured on CIM medium composed of 1x Gamborg's B-5 salts, 20 g/l glucose, 0.5 g/l MES, 1x Gamborg's vitamin solution and 1% Phytoagar supplemented with 2,4-D (500 µg/l) and kinetin (50 µg/l). CIM calli appeared after 7 to 10 days and were propagated in two weeks-subculture intervals on CIM medium. To generate the callus line with an egg cell-like expression profile, the coding sequence of RKD2 (AT1G74480) was amplified from pistil cDNA using the primer pair RKD2fw and RKD2rev (Supplementary Data 11). Pistils from flowers at developmental stage 12 [47] were harvested to purify mRNA and generate cDNA as described [48]. The PCR fragment was cloned into pENTR/D-TOPO (Invitrogen, Carlsbad, USA) and subsequently transferred into the GATEWAY-compatible destination vector pH7FWG2.0 [49] with LR clonase. The resulting expression vector 35Sp:RKD2-GFP was used for floral dip transformation of Arabidopsis [50]. Seeds of transformed plants were surface-sterilized and grown on ½ MS medium with 2% (w/v) sucrose, 1% Phytoagar, and Hygromycin (30 µg/ml). RKD2-induced calli had formed after 20 to 30 days and were propagated in two weeks-subculture

intervals on ½ MS, 2% (w/v) sucrose, 1% Phytoagar, Hygromycin (30 µg/ml). Calli were collected with a sterile scalpel blade and immediately frozen in liquid nitrogen.

***Leaf and root samples:*** Arabidopsis plants (Col-0) were grown under continuous light conditions at 22°C. Senescent rosette leaves were collected from 35 days-old plants. Samples for stem, first node, second internode and first cauline leaf were collected from 30 days-old plants. Rosette leaf sections (Leaf7: distal; proximal; petiole) were harvested from 22 days-old plants before bolting. Seedlings were grown on ½ MS plates under continuous light for seven days and separated into cotyledons, hypocotyl, root tip, root maturation zone or seed apical meristem including cotyledons and first leaves. Whole roots were harvested from 22 days-old plants grown under continuous light on ½ MS plates.

Classification of growth stage and plant section was done as described [44,51-53]. Harvested material from at least three individual plants was combined for each sample, frozen in liquid nitrogen and stored at -80°C until further use.

***Protein lysis and digest***

Frozen plant material was homogenized with a tissue lyzer (Qiagen, Hilden, Germany) or with mortar and pestle in liquid nitrogen. Proteins were precipitated over night with 10% tricholoroacetic acid in acetone at -20° C and subsequently washed two times with ice-cold acetone. Dry samples were incubated with urea digestion buffer (8 M urea, 50 mM Tris-HCl pH 7.5, 1 mM DTT, cOmplete™ EDTA-free protease inhibitor cocktail (PIC) [Roche, Basel, Switzerland], Phosphatase inhibitor [PI-III; in-house, composition resembling Phosphatase inhibitor cocktail 1,2 and 3 from Sigma-Aldrich, St. Louis, USA]) for 1 h. Protein concentration was determined with a Bradford assay [54]. 1 mg of protein was reduced (10 mM DTT, 1h, room temperature), alkylated (55 mM chloroacetamide, 30min, room temperature) and subsequently diluted 1:8 with digestion buffer (50 mM Tris-HCl pH 8.0, 1 mM $CaCl_2$). In-solution pre-digestion with trypsin (Roche, Basel, Switzerland) was performed for 4 hours at 37°C (1:100 protease:protein ratio), followed by overnight digestion with trypsin (1:100 protease:protein ratio). Samples were acidified to pH 3 using trifluoroacetic acid (TFA) and centrifuged at 14,000 *g* for 15 min at 4°C. The supernatant was desalted on 50 mg SepPAC SPE catridges (Waters, Milford, USA). Peptides were eluted with 0.1 % TFA in 50% acetonitrile (ACN) and vacuum-dried in a Thermo Savant SPD SpeedVac (Thermo Fisher Scientific, Waltham, USA).

***Peptide enrichment and off-line fractionation***

$Fe^{3+}$-IMAC was performed as described previously with some adjustments [55]. Briefly, desalted peptide samples were re-suspended in ice-cold IMAC loading buffer (0.1% TFA, 40% ACN). For quality control, 1.5 nmol of a synthetic library of phosphopeptides and their corresponding non-phosphorylated counterpart sequence (B2 and F1) [56] were spiked into each sample prior to loading onto a $Fe^{3+}$-IMAC column (Propac IMAC-10 4x50 mm, Thermo Fisher Scientific, Waltham, USA). The enrichment was performed with Buffer A (0.07% TFA, 30% ACN) as wash buffer and Buffer B (0.315% $NH_4OH$) as elution buffer. Collected full proteome and phosphopeptide fractions were vacuum-dried and stored at -80°C until further use.

For the full proteome fraction, hydrophilic strong anion exchange chromatography (hSAX) peptide separation was performed as described previously [15]. Briefly, an equivalent of 300 µg protein digest was reconstituted in hSAX solvent A (5 mM Tris-HCl, pH 8.5) and separated using a Dionex Ultimate 3000 HPLC system (Dionex Cor., Idstein, Germany) equipped with an IonPac AG24 guard column (2x50 mm) and an IonPac AS24 stong anion exchange column (2x 250 mm, Thermo Fisher Scientific, Waltham, USA). Fractions were collected in 96 well format and subsequently pooled to 24. Individual fractions were acidified with formic acid (FA), desalted on self-packed StageTips (five disks, Ø 1.5 mm C18 material, 3M Empore™, elution solvent 0.1% FA in 50% ACN) and dried down prior to LC-MS/MS analysis. Phosphopeptide fractions were separated into four fractions using a StageTip (five disks, Ø 1.5 mm C18 material, 3M Empore™) based high pH reversed phase protocol as described previously [57]. Phophosphopeptides were eluted with 2.5%, 7.5%, 12.5% and 50% ACN in 25 mM $NH_4FA$ pH 10. The flow through and 50% ACN fraction were combined and all fractions were dried down prior to LC-MS/MS analysis.

For the cycloheximide and MG132 chase experiments, 300 µg peptides for each sample were reconstituted in high pH reversed phase loading buffer (2.5 mM $NH_4HCO_3$, pH 8) and fractionated using a Dionex Ultimate 3000 HPLC system (Dionex Corp., Idstein, Germany) and a Waters XBridge column (BEH130 C18, 3.5 um, 2.1x150mm; Waters, Milford, USA). Peptides were separated by a linear gradient from 4% Buffer D to 32% D in 45 min, followed by a linear gradient from 32% D to 80% D in 6 min. The proportion of Buffer A was kept at 10% during fractionation (Buffer A: 25 mM $NH_4HCO_3$, pH 8; Buffer C: 100% ACN; Buffer D: 100% ultrapure water). Fractions were collected in 96 well format, subsequently pooled to 48, acidified and dried in a SpeedVac.

***Seed treatment with cycloheximide and MG132***

Aliquots of surface-sterilized mature dry seeds ((10 mg, Col-0) were stored for two days at 4°C in the dark and subsequently imbibed with 2 ml liquid ½ MS at 22°C under constant light. After 4 h incubation, cycloheximide (CHX) or MG132 (*N*-(benzyloxycarbonyl)-Leu-Leu-Leu-al) was added to a concentration of 100 µM. Since both CHX and MG132 were dissolved at the same stock concentration in 100% dimethylsufloxide (DMSO), the respective volume of 100% DMSO was added to the control samples. As a baseline sample, one seed aliquot was dried and immediately frozen in liquid nitrogen after the 4 h incubation in ½ MS medium. Seeds for CHX, MG132 and DMSO control were removed after 8 h, 16 h and 24 h treatment, dried and frozen in liquid nitrogen. Germination of treated seeds was visually checked after 4 days of incubation. Both compounds CHX and MG132 were active and led to complete or partial inhibition of seed germination as compared to a DMSO treated control sample (Extended Data Fig. 4i).

Protein extraction and digest was performed as described above. Peptide quantification prior to high pH reversed phase fractionation was done using the Pierce™ BCA protein assay kit (Thermo Fisher Scientific, Waltham, USA) [58].

***Size exclusion chromatography (SEC)***

Homogenized flower (stage 15), leaf (rosette leaf 7) and root (whole root) samples were lysed in ice-cold 50 mM Tris-HCl pH 7.4, 100 mM NaCl, 10% glycerin, PIC, PI-III and 0.1% Triton-X100. After filtering (Ø 0.2µm), 0.25 ml lysate containing 1 mg of protein was injected on a Superose 6 10/30 GL column (GE Healthcare, Chicago, USA) and separated at a flow rate of 250 µl/min on an Äkta pure 25 (GE Healthcare, Chicago, USA). Molecular weight calibration of the column was performed with the high molecular weight gel filtration calibration kit (GE Healthcare, Chicago, USA). After the void volume, 80 fractions of 125 µl each were collected, vacuum dried and re-solubilized in urea digestion buffer. In-solution digestion with trypsin and sample desalting on self-packed StageTips was performed as described above. For quality control, PROCAL peptide standard [59] was spiked into each SEC fraction prior to LC-MS/MS analysis.

***LC-MS/MS analysis***

Nanoflow LC-MS/MS was performed by coupling a Dionex 3000 (Thermo Fisher Scientific, Waltham, USA) to a QExactive Orbitrap HF (Thermo Fisher Scientific, Waltham, USA). Samples for the proteome and phosphoproteome analysis were re-suspended in loading buffer containing 0.1% formic acid (FA) or 50 mM citrate and 1% FA, respectively. Peptide loading and washing were done on a trap column (100 µm i.d. x 2 cm, packed in-house with Reprosil-Pur C18-GOLD, 5 µm resin, Dr. Maisch, Ammerbuch, Germany) at a flow rate of 5 µl/min in 100%

loading buffer (0.1% FA) for 10 min. Peptide separation was performed on an analytical column (75 µm i.d. x 40 cm packed in-house with Reprosil-Pur C18, 3 µm resin, Dr. Maisch, Ammerbuch, Germany) at a flow rate of 300 nl/min using a 110 min gradient from 4% to 32% solvent B (solvent A: 0.1% FA, 5% DMSO in HPLC grade water; solvent B: 0.1% FA, 5% DMSO in acetonitrile) for the tissue map full proteome analysis and a two-step 110 min gradient from 2% to 27% solvent B for the phosphoproteome analysis [60]. Peptides were ionized using a spray voltage of 2.2 kV and a capillary temperature of 275°C. The instrument was operated in data-dependent mode, automatically switching between MS and MS2 scans. For the full proteome samples of the tissue atlas experiment, full scan MS spectra (m/z 360 – 1300) were acquired with a maximum injection time of 10 ms at 60,000 resolution and an automatic gain control (AGC) target value of 3e6 charges. For the top 12 precursor ions, high resolution MS2 spectra were acquired in the orbitrap with a maximum injection time of 50 ms at 15,000 resolution (isolation window 1.7 m/z), an AGC target value of 2e5 and normalized collision energy of 25%. The underfill ratio was set to 1% with a dynamic exclusion of 30 s. Only precursors with charge states between 2 and 6 were selected for fragmentation. For the phosphoproteome analysis, the MS2 spectra were acquired with a resolution of 30,000 (isolation window 1.7 m/z) and a maximum injection time of 120 ms. Dynamic exclusion was set to 35 s.

SEC samples were measured using a 60 min 4% to 32% gradient. MS2 spectra were generated for the top 20 precursors, with a maximum injection time of 50 ms at 30,000 resolution (isolation window 1.7 m/z), a normalized collision energy of 25% and a dynamic exclusion of 20 s.

The samples for cycloheximide and MG132 treated seed samples were analyzed by a micro-flow LC system coupled online to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Waltham, USA)) as described previously [61]. The micro LC system was built by combining a modified Vanquish pump with the auto-sampler of the Dionex UltiMate 3000 RSLCnano System. The sample was directly loaded onto a Thermo Fisher Scientific Acclaim PepMap 100 C18 LC column (2 µm particle size, 1 mm ID x 150 mm), the flow rate was 50 µl/min and column temperature was maintained at 55 °C. A 15 min linear gradient of 7%-32% solvent B (solvent A: 0.1% FA, 3% DMSO in HPLC grade water; solvent B: 0.1% FA, 3% DMSO in acetonitrile) was used to separate all the samples, followed by 1 min 95% solvent B washing and 1 min 0.5% solvent B equilibrium time at a flow rate of 100 µl/min. The Orbitrap Fusion Lumos mass spectrometer was operated with the Ion Max API Source installed with a HESI-II probe (50 µm ID). The detailed acquisition parameters are: Positive polarity; spray voltage 3.5 kV, funnel RF lens value at 40, capillary temperature of 325 °C, auxiliary gas heater

temperature of 300 °C. The flow rates for sheath gas, aux gas and sweep gas were set to 32, 5, and 0, respectively. Full scan MS spectra (m/z 360 – 1300) were acquired in the orbitrap with a maximum injection time of 10 ms at 120,000 and an AGC target of 4E5. MS2 spectra were acquired in the linear ion trap (rapid scan mode) after collision-induced dissociation (CID) fragmentation with collision energy set at 35%, an AGC target of 1E4 and maximum IT of 10 ms. The isolation window was set to 0.4 m/z and scans were recorded with a maximum duty cycle of 0.6 s and the option 'inject ions for all parallelizable time' enabled. Only precursors with charge states between 2 and 6 were selected for fragmentation and the lowest scan range of MS2 was fixed at 100 m/z. The intensity threshold was set to 5E3 and the dynamic exclusion to 12 s.

***Peptide and protein identification and quantification***

Raw data files were searched with MaxQuant software (v. 1.5.8.3) using standard settings unless otherwise described [62]. MS/MS spectra were searched against Araport11 [5] protein coding genes (Araport11_genes.201606.pep.fasta; download 06/2016), with trypsin as protease and up to two allowed missed cleavages. Carbamidomethylation of cysteines was set as fixed modification and oxidation of methionines and N-terminal acetylation as variable modifications. The 'match-between-runs' function was enabled for corresponding fractions within one parameter set, where applicable.

For the tissue map analysis, full proteome and phosphoproteome samples were processed together as two separate parameter groups, with phosphorylation of serine, threonine or tyrosine defined as variable modification only for the phosphoproteome group. Here we also added the spike-in phosphopeptide library sequences [56] to the database search space.

For sORF identification two sORF databases (ATSO and ARA-PEP) were added to the search space in addition to Araport11. ARA-PEP is a previously described repository of putative sORF-encoded peptides in Arabidopsis [14]. ATSO (*Arabidopsis thaliana* sORFs) was generated by using sORFfinder [63] and our RNA-seq analysis data to identify putative new sORFs in non-coding intergenic regions (ATSO.non_coding.pep.nr.cd-hit_aS1_aL0.3_c1.out.nr.fasta). Three targets were used as input for sORFfinder. Sequences of intergenic and non-coding regions relative to the Araport11 annotation and sequences resulting from a Trinity [64] *de novo* transcriptome assembly which are non-overlapping with the Araport11 annotations. Afterwards, the sORF nucleotide sequences were translated to amino acid sequences. CD-HIT [65] was used with the parameters –aS 1 –aL 0.3 –c 1 to reduce sequence redundancy. .

SEC experiment raw files for flower, leaf and root were all searched together, with each SEC fraction designated as individual experiment and the 'match-between-runs' function enabled.

The seed experiment raw files were searched using standard settings as described above.

Peak list files generated by Baerenfaller et al. (2008) [7] were downloaded from Pride (PRD000044) and searched with the Mascot [66] search engine (version 2.4.1, Matrix science) against the Araport11 database. The target-decoy option of Mascot was enabled and search parameters included a precursor tolerance of 1.3 Da and a fragment ion tolerance 0.45 Da. Enzyme specificity was set to trypsin and up to two missed cleavages were allowed. The Mascot $^{13}C$ option, which accounts for the misassignment of the monoisotopic precursor peak, was set to 1 and oxidation of methionine and carbamidomethylation of cysteines were included as variable and fixed modification, respectively. The isobarQuant workflow was used to generate protein identification files [67].

***Spectra validation***

sORF peptides identified by database searching that could not be mapped to an existing Araport11 gene model were synthethized at JPT Peptide Technologies (Berlin, Germany) using Fmoc-based SPOT synthesis on membranes [59] and measured on the same mass spectrometer that was used for data acquisition (Tissue atlas full proteome method; see above for method description). Experimental and synthetic peptide spectra were extracted from the raw files and used for similarity calculation without prior processing. Normalized spectral contrast angle (SA) comparison between spectra of the tissue sample and synthetic peptides was performed using in-house R scripts [68] factoring in peaks which are either shared between spectra or exclusive to the synthetic peptide spectra. Peaks exclusive to experimental spectra were ignored. Candidates were selected with SA angle cutoff larger than 0.7 and BLASTed against the UniProt database for further validation (Supplementary Data 3).

Protein identifications for the 'uncertain' evidence category of UniProt were validated by comparing the experimental spectra to in-silico predicted fragment spectra (Supplementary Table 1) [12]. For this, a spectral library was obtained from Prosit (https://www.proteomicsDB.org/prosit) [12] by uploading all sequence-charge combinations of peptides that were identified for 'uncertain' proteins. To obtain the best matching spectra, the collision energy of Prosit was calibrated using a standard quality control run from the same mass spectrometer that was used for data acquisition. The resulting predicted spectral library

was visualized and compared to the experimentally acquired data analog to the sORF candidate peptides. SA values larger than 0.7 were used as cutoff.

***RNA sequencing***

Total RNA was isolated with the NucleoSpin RNA Plant kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. DNA was removed by on-column treatment with rDNAse (Macherey-Nagel, Düren, Germany). For recalcitrant samples (seed, silique, pollen), a LiCl-based protocol was adopted with minor modifications [69]. After LiCl precipitation, the RNA pellet was dissolved in rDNAse buffer and treated with rDNAse (Macherey-Nagel, Düren, Germany) at 37°C for 10 min. The final pellet was re-suspended in 35 µl DEPC-treated water.

RNA was quantified (Nanodrop™, Thermo Fisher Scientific, Waltham, USA ) and quality checked with a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA). RNA integrity number (RIN) values between 6.4 and 10 were accepted for further analysis. cDNA libraries were prepared using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, San Diego, USA) according to the instructions. Clusters were generated in two batches and sequenced on a High throughput flow cell with the HiSeq 2500 platform (Illumina, San Diego, USA) to a depth of 36 million reads per sample. Quality assessment of raw and trimmed 75 bp paired RNA-seq reads was performed with FastQC. Raw RNA-seq reads were trimmed to remove adapter contamination and poor quality base calls using Trimmomatic version 0.35 with parameters (ILLUMINACLIP: Illumina-PE.fasta:2:30:10; LEADING:3; TRAILING:3; SLIDINGWINDOW:4:20; MINLEN:36) [70]. Trimmed RNA-seq reads were mapped to the Araport11 [4] transcriptome with Kallisto version 0.43.1 (default parameters) [71].

***Data processing***

MaxQuant output tables were filtered for non-plant contaminants, reversed sequences and proteins which were only identified based on modified peptides. Protein abundance estimation was based on either intensity-based absolute quantification (iBAQ) [26], or top3 quantification [72], depending on the analysis. MaxQuant ProteinGroups containing several gene loci were filtered out in order to retain only unambiguously identified gene loci for further analyses. In case multiple protein isoforms were identification in distinct ProteinGroups, only the isoform with the higher number of razor+unique peptides was retained.

mRNA quantities are displayed as transcripts-per-kilobase-million (TPM) and a cutoff of 1 TPM was used as lower limit of detection across all samples. Gene ontology [73] (GO)-term analysis for the transcript abundance range was performed using the 1D enrichment function from Perseus

[74,75] with Benjamini-Hochberg FDR threshold set to 0.01. In this function, a two-sided Wilcoxon-Mann-Whitney test is employed to test for systematically larger or smaller transcript abundance within a GO category as compared to the global distribution of all values (Supplementary Data 4).

For further qualitative and quantitative analyses, all transcript or protein isoform information was collapsed onto gene level. Unless otherwise stated, displayed abundances for protein, transcript and phosphorylation sites (p-site) were $\log_2$ transformed. Protein, peptide and transcript datasets for the tissue map and seed treatment experiments were median centered to the overall median of the respective dataset. No normalization was performed for the p-site dataset, since total p-site intensity variations between tissues are also due to biological sample differences. Instead, the spike-in phosphopeptide library was used, to assess reproducible enrichment efficiency and MS measurement quality of phosphoproteome samples [56].

Phosphorylation sites (p-sites) which were reported in the MaxQuant output table (Phospho(STY)sites.txt file) with a 'localization probability' larger than 0.75 were designated as 'high confidence localizations' or class I sites [76]. P-sites were considered exclusive if they were only detected in a single tissue. The number and identity of p-sites and phosphoproteins detected in this study were compared to datasets available through PhosPhAT4.0 and a published meta-analysis [8,9] (Extended Data Fig. 1h).

***Data analysis***

***Genome annotations:*** Chromosome information contained in the Arabidopsis locus identifiers (AGI codes: AT [Arabidopsis thaliana]; 1, 2, 3, 4, 5, M, C [chromosome number, M for mitochondrial, C for chloroplast]; G [gene]; 12300 [five-digit code for position on chromosome]), was used to assign genes to their respective chromosomes (Extended Data Fig.1c). Araport11 gene identifiers (Arabidopsis gene identifiers [AGI]) were mapped to the UniProt *Arabidopsis thaliana* reference proteome (taxon identifier 3702; UP000006548; download 11/2018) based on protein sequence. Swiss-Prot, TremBL as well as protein existence criteria annotations (level1: evidence on protein level; level2: evidence on transcript level; level 3: inferred from homology, level 4: predicted, level 5: uncertain) were subsequently obtained from UniProt (Extended Data Fig.1d).

N- and C-terminal peptide sequences were extracted from the MaxQuant peptides.txt file and filtered for zero missed cleavages (n = 2,776) (Extended Data Fig. 2a). N-terminal peptides were divided into groups with (n = 1,707) or without (n = 1,069) cleavage of the initiator

methionine. The frequency of the 20 genetically encoded amino acids at the position after the start codon was subsequently calculated separately for both groups and displayed as a pie chart. For N-terminal peptides with the same amino acids at the position after the start codon, the percentage of peptides with N-terminal acetylation was also calculated and displayed as bar plot. This analysis was also performed separately for the groups with or without cleavage of the initiator methionine (Extended Data Fig. 2b,c).

Gene isoforms annotated in Araport11 were considered as 'identified', if they were detected with a TPM intensity larger than our limit of detection cutoff (1 TPM; transcript level) or with an isoform-specific peptide (protein level) in at least one tissue sample (Extended Data Fig. 2f, Supplementary Data 3). A selection of isoform-specific peptides was synthesized together with the sORF peptide collection at JPT Peptide Technologies (Berlin, Germany) and the synthetic peptides were used for validation of the experimental spectra identifications as described above.

Transcription factor, transcription regulator, kinase and phosphatase family annotations and classifications are based on reports by Zheng et al. [16] and Yang et al. [17]. Proportional coverage of these families within our dataset was calculated by counting how many of them could be identified on transcript, protein or phosphoprotein level, respectively (Extended Data Fig. 2k).

***Tissue groups and tissue characteristics:*** Tissue samples were assigned to tissue groups as follows, based on their origin or common morphology: flower (sepal, petal, stamen, carpel, silique, flower); fruit (silique septum, silique valves); seed (embryo, seed, seed imbibed); pollen (pollen); leaf (cauline leaf, leaf distal, leaf proximal, leaf petiole, senescent leaf, cotyledons, shoot tip); stem (node, internode, flower pedicle, hypocotyl); root (root, root tip, root upper zone); callus (egg-cell like callus, callus); cell culture (cell culture early, cell culture late) (Fig. 1a,b).

To provide a measure of tissue similarity, the Pearson correlation coefficient was calculated for the gene expression of all pair-wise tissue combinations. The Pearson correlation coefficient is a measure of the linear correlation between two variables, in this case the expression levels in two tissue samples. Correlations were computed both on protein level and transcript level and displayed as separate heatmaps for both omics levels (Extended Data Fig. 1a). For three examples of morphologically highly similar tissue pairs, the gene expression levels were also displayed as scatter plots using protein abundance or transcript abundance measurements, respectively (Extended Data Fig. 1b).

Tissue or tissue group specificity of genes was calculated based on iBAQ or TPM abundance values, respectively [77]. Genes were assigned to categories in the following order: 'tissue specific', 'tissue enhanced', 'group-specific', 'group-enhanced', 'shared' and 'mixed'. Genes were considered 'specific', if they were only detected in a particular tissue or tissue-group and 'enhanced' if their abundance was at least five-fold higher in a particular tissue or tissue-group as compared to the average levels in all other tissues. Genes that were detected in all 30 tissues but did not show enhanced expression in a tissue or tissue group were classified as 'shared'. All remaining genes are contained in the 'mixed' category (Extended Data Fig. 3a).

Genes detected in all tissues on either protein or transcript level were assigned to the core datasets (core_transcript n= 8,405; core_protein n = 7,734; core_intersection n = 5,043). This classification does not consider gene expression levels and should not be confused with the tissue-specificity classification described in the previous paragraph.

Hierarchical clustering analysis for tissues representing flower parts or the whole flower was performed on $\log_2$-transformed, z-scored protein intensities (iBAQ values) in Perseus using Euclidean distance and average linkage (Extended Data Fig. 3b). The flower organ identity model was restricted to the simplified ABC model, thus only displaying homeotic genes for the A, B and C classes [78], namely the expression of MADS-box transcription factors *APETALA1* (*AP1*) for class A, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) for class B and *AGAMOUS* (*AG*) for class C. Class A expression specifies sepal formation, the combination of class A and B specifies petal formation, the combination of class B and C specifies stamen formation and class C expression specifies carpel formation in developing flowers.

For the cumulative abundance calculation of five representative tissues (flower; pollen; root tip; leaf proximal; seed), proteins (iBAQ) and transcripts (TPM) were first ranked from highest to lowest individual intensity (Extended Data Fig. 3e). The running total was then plotted against the rank order. The names or identifiers of the five most abundant transcripts or proteins (rank 1 to 5) are listed in descending order for the respective tissue. Shared IDs among the top 100 most abundant transcripts (TPM) and proteins (iBAQ) were calculated for each individual tissue (Extended Data Fig. 3f). For the genes representing the highest abundant protein in at least one tissue, the contribution to the total protein amount in an individual tissue was calculated by dividing the iBAQ intensity (not log transformed) of the respective protein by the summed total iBAQ intensity (not log transformed) in this tissue (Extended Data Fig. 3g).

Principal component analysis (PCA) of proteome and transcriptome data was performed in Perseus for the intersection of the core datasets (n = 5,043) on $\log_2$-transformed and z-scored

iBAQ and TPM intensities (Extended Data Fig. 3h). Subcellular localization information was downloaded from SUBA (download 11/2016; unambiguous localizations n = 3,506) [44] (Supplementary Data 2) and used to calculate the intensity contribution of proteins assigned to a specific compartment for the different tissue groups. For this, the percentage of proteins with the same SUBA annotation was calculated by dividing the summed protein intensity (iBAQ, not log transformed) of each SUBA category by the total summed iBAQ intensity (not log transformed) within each tissue. To compare tissue groups, the respective protein intensity proportion was then averaged for all tissues within one group. These averaged proportions were subsequently plotted for each subcellular compartment (Extended Data Fig. 3i).

***Protein/mRNA relation*:** The Pearson correlation value between median protein (median iBAQ of 30 tissues) and median transcript (median TPM of 30 tissues) abundances was calculated using the set of genes with abundance measurements on both protein and transcript level in more than ten tissues (more than 10 pairwise complete observations; n = 14,069). The scatterplot of median transcript (TPM) and protein abundance (iBAQ) was displayed together with their marginal histograms (Fig. 2a). Tissue-specific Pearson correlation values between protein and transcript abundance were calculated using all genes from the core proteome and core transcriptome intersection dataset (n = 5,043; Extended Data Fig. 4a). The core dataset was used here to base the Pearson correlation value comparison between tissues on the same set of genes in all tissues.

The data subset with more than 10 pairwise complete observations on protein and transcript level (n = 14,069) was also used for all further calculations involving protein-to-mRNA ratios (PTR) (Fig. 3d, Extended Data Fig. 5). PTR values are calculated by building the ratio between protein abundance (iBAQ) and the corresponding transcript abundance (TPM) for each gene and were calculated separately for each tissue. The tissue PTR values were then used to calculate the median and median absolute deviation (MAD) of PTR values across all 30 tissues (Extended Data Fig. 5). The variation of PTR values across tissues indicates, whether protein and transcript levels of a given gene are regulated in a similar (small PTR variation) or different way (high PTR variation) between individual tissues. To estimate the proportion of genes with rather stable or variable PTR values, the PTR MAD values were distributed into five equal parts, so that 20% of all genes from this analysis are contained in one part (5 MAD quantiles: Q1-Q5). For each gene, the median PTR values were plotted against their MAD across tissues and the MAD quantiles indicated in the plot (Extended Data Fig. 5a). GO-term analysis for the median PTR distribution (median PTR across all tissues) was performed using the 1D enrichment

function from Perseus [74] with Benjamini-Hochberg FDR threshold set to 0.01 (Supplementary Data 4). In this function, a two-sided Wilcoxon-Mann-Whitney test is employed to test for systematically larger or smaller PTR values within a GO category as compared to the global distribution of all values.

In a similar way, the variation of phospho ratios across tissues indicates, whether p-site levels are driven by protein abundance (small phospho ratio variation) or p-site stoichiometry (high phospho ratio variation) changes, respectively. Phospho ratios were calculated by building the ratio between a p-site abundance (intensity) and the abundance of its corresponding protein (iBAQ). For this, we used the set of p-sites, with abundance measurements on both p-site and protein level in more than ten tissues (more than 10 pairwise complete observations; n = 13,793). Phospho ratios were calculated separately for each tissue and subsequently combined to calculate the median phospho ratio and phospho ratio MAD across all 30 tissues. To estimate the proportion of p-sites which closely resemble the protein abundance profile, the phospho ratio MAD values were again distributed into five equal parts (5 MAD quantiles: Q1-Q5), and the median phospho ratio values plotted against their MAD across tissues each p-site. MAD quantiles were indicated in the plot (Extended Data Fig. 5c).

To provide further information about inter-tissue variability, MAD quantiles were also calculated and displayed for the expression levels on protein (iBAQ; n = 14,069), transcript (TPM; n = 14,069) and p-site (peptide intensity; n = 13,793) levels, respectively (Extended Data Fig. 5b,d).

***Features for protein level prediction models*:** Since a substantial proportion of variation in protein levels remained unexplained when using transcript level information alone (Pearson correlation between protein and transcript abundance in tissues 0.28-0.79), additional molecular features, that could explain protein abundance variations, were tested for their influence on protein abundance in a model selection approach. Predictors selected for this analysis were: mRNA levels, codon usage, non-synonymous to synonymous substitutions (*Dn/Ds*) ratios, which are a measure of evolutionary conservation, gene/coding sequence (CDS) length, exon number, gene position on the chromosome, cytosine methylation, the number of putative protein interactions and mRNA sequence motifs (kmers of size 3 to 7 nucleotides):

Codon usage statistics for the *A. thaliana* genome were obtained from Kazusa (www.kazusa.or.jp/codon/current/species/3702) and parsed to extract NCBI gene identifiers. These identifiers were mapped to their corresponding UniProt entries and Ensembl/TAIR10 annotations using the UniProt Retrieve/ID mapping tool. The extracted TAIR10 annotation was

merged with the Kazusa codon usage dataset. Codon frequencies were calculated for each gene by dividing the count (x 3) of a given codon by the full length of the coding sequence.

The *Dn/Ds* substitution rates were calculated from CDS pairs between *Arabidopsis thaliana* and its closest relative *Arabidopsis lyrata*. Reciprocal best BLAST with an e-value cutoff of ≤ 1E-08 was used to identify ortholog sequences. Individual CDS pairs were aligned using PRANK [79] and Gblocks [80] was applied to eliminate poorly aligned positions in an alignment with a cutoff of 8 contiguous non-conserved positions and without allowing for gap positions. The yn00 package in the program PAML [81] for pairwise sequence comparison was used to estimate substitution rates, *Dn* and *Ds*, respectively.

The total number of exons and the total gene lengths were obtained from Araport11. The distance of each gene from the chromosomal centromeres were calculated to capture potential position-specific effects.

*A. thaliana* (Col-0) Whole Genome Bisulphite Sequencing (WGBS) data was obtained from van der Graaf et al. 2015 [82]. For each gene, methylation levels were calculated for contexts CG, CHG and CHG (where H = adenine, cytosine, thymine) separately. Per gene methylation levels were defined as:

$$g_i = \frac{1}{max(j)} x \sum \frac{Nm_j}{N_j}$$

Here, *max(j)* is the total number of cytosines, $Nm_j$ is the number of methylated reads and $N_j$ the total number of reads at the jth cytosine.

Arabidopsis protein-protein interactions were downloaded from STRING [30] and the number of protein interaction partners extracted for each gene.

All of these features have previously been associated with transcriptional activity and/or protein abundance levels in Arabidopsis and/or other species [83-88]. For a detailed feature set description see Supplementary Data 5.

***De novo motif identification*:** Motifs in the mRNA CDS, 3' or 5'UTR region were identified as previously described [23]. Briefly, protein expression levels were $\log_{10}$ transformed and median centered. For genes with two or more transcript isoforms, the transcript isoform reported to have the largest summed iBAQ values across all tissues was defined as the major transcript isoform per gene and used to compute all sequence feature and mRNA levels. Raw RNA-seq reads were trimmed to remove adapter contaminations and poor quality base calls using Trimmomatic

0.39 with parameters (ILLUMINACLIP:Illumina-PE.fasta:2:30:10; LEADING:3; TRAILING:3; SLIDINGWINDOW:4:20; MINLEN:36). Trimmed RNA-seq reads were mapped to the Araport11 transcriptome annotation with Kallisto version 0.46.0 using default parameters. To estimate the levels of mature mRNA, the number of reads mapping to exonic and intronic regions of the transcripwere counted separately and then normalized by the total exonic and intronic region lengths, respectively. Normalized intronic counts were subtracted from the normalized exonic counts to obtain the mature mRNA counts. The resulting normalized exonic counts per sample were corrected by the DESeq2 [89] library size factor and $\log_{10}$ transformed. Transcripts with 10 reads per 1kb were treated as transcribed. Tissue-specific protein-to-transcript ratios (PTR) were computed using the normalized protein and transcript levels. GEMMA software [90] was used to identify de-novo motifs in 5'UTR, CDS and 3'UTR regions using the tissue-specific PTRs as response variable. GEMMA uses a linear mixed model, in which the effect of each individual kmer on the median PTR across tissues is assessed while controlling for the effect of other kmers (random effects), region length and GC percent (fixed effects). The motif search was performed for kmers ranging from 3-7 nucleotides. Obtained p-values were adjusted for multiple testing with Benjamini-Hochberg's false discovery rate (FDR) and jointly computed across the p-values of all tissues. Gemma was run using the median PTR with FDR<0.1 and covariates set to 'false'. 82 significant putative motifs were obtained based on their sequence (5'UTR n = 32; 3'UTR n = 38; CDS n = 12) and sub-sequence (initial, all, end) region. 39 motifs that lie in the 'initial' and 'end' sub-sequences were further selected. The presence or absence of each enriched motif with respect to each gene was extracted in form of a binary matrix and used for downstream multivariate feature selection analysis.

***Model-based feature selection*:** Tissue-specific protein expression data was merged to the feature matrix. Preliminary pair-wise correlation analysis showed only weak to moderate correlation between individual features. In addition, Variance Inflation Factors (VIF) were calculated for each feature using the fmsb R package [91]. The result showed low VIFs, suggesting that multicollinearity was not an issue for downstream analyses.

Two model selection approaches were implemented, stepwise regression and Lasso regression [92,93]. To select the most predictive features for protein abundance in each tissue, we used a forward-backward model selection approach in a multiple regression framework. The method was implemented using stepwiseAIC() function in R [94], which compares the fit of nested models. To ensure that the comparison of model AICs were not affected by unequal sample size, missing data were removed prior to the analysis.

Because stepwise regression can occasionally lead to over-fitting [95], a Lasso regression was implemented as an alternative model selection procedure. Lasso regression performs L1 regularization, which adds a penalty equivalent to the absolute of the magnitude of regression coefficients and tries to minimize them. The strength of the penalty was controlled via the tuning parameter λ [93]. Lasso was implemented using the glmnet package [96] in R. Model training was performed on a random selection of genes (50% of the dataset) and implemented using the cv.glmnet() function. The optimal value for λ was extracted and used to re-build the model using the glmnet() function. Finally, the fitted model was used to obtain predictions in the remaining 50% of the genes.

For each tissue, stepwise and Lasso regression models were compared. Stepwise regression yielded only slightly higher coefficients of determination ($R^2$) values compared to Lasso, suggesting over-fitting was not an issue (Supplementary Data 5). As stepwise regression yielded more parsimonious models in general, we used this approach for further analysis. Selected features here could explain on average 52% of protein abundance variation (37% in pollen, 62% in cell culture early) (Supplementary Data 5). For each tissue, features from the best fitting models were summarized in an incidence matrix along with the effect direction (positive or negative effect on protein levels). To determine the importance of each feature to the overall model fits, $R^2$ variance decomposition was performed using the 'genici' metric which is implemented in the relaimpo R package [97] (Extended Data Fig. 4c). Relative feature contributions were averaged across all tissues (Fig. 3b).

Since many of the detected motifs appear in a tissue-specific manner (predictive only in tissue subset), we clustered tissues according to the presence or absence status of 5'UTR motifs in each tissue model (Extended Data Fig. 4d). Seed and pollen tissues form a distinct subcluster, which might be connected to the increased/decreased PTR levels in these tissues compared to the other tissues (Fig. 3c;Extended Data Fig. 4a). In contrast to the motif analysis, Dn/Ds ratios were consistently (and positively) associated with protein abundance in all 30 tissues (Supplementary Data 5). To explore the relationship between Dn/Ds ratios and protein abundance in more detail, groups of genes with low or high (bottom 5% or top 5% of Dn/Ds distribution) were compared to each other with regard to protein levels (Extended Data Fig. 4e).

***Seed tissue PTR regulation*:** The PTR value distribution was plotted for the median PTR (median across all tissues, see above) and selected tissues (seed, seed imbibed, pollen, cell culture young). Seed and pollen tissues show particularly low Pearson correlation between protein and transcript abundance levels (Extended Data Fig. 4a). Cell culture early in

comparison shows the highest Pearson correlation among the analyzed tissues. Tissue-specific 'outliers' with especially high or low PTR as compared to all other tissues were selected if their PTR value was outside two standard deviations of the mean of the median PTR distribution. For seed, this resulted in 469 genes with 'high' and 571 genes with 'low' PTR. In order to interpret, whether this change in seed PTR values was caused by differential regulation on transcript and/or protein level, the fold change between seed protein and transcript abundance and the median protein and transcript abundance of all tissues was calculated. Genes with at least 4fold-change in either protein or transcript abundance were considered regulated (less/more transcript and/or less/more protein in seed compared to all tissues). The percentage of genes with high or low PTR levels in the seed tissue, which are regulated in the same manner (less/more transcript and/or protein) was calculated and displayed as arrow plot (Fig. 3e). About a quarter of each gene set (high PTR 27%; low PTR 24%) show no regulation based on our 4fold-change cutoff.

For the low PTR gene set, a high proportion of genes with lower protein abundance in the seed tissue was observed. A protein level chase experiment was performed, to investigate whether this reduction was due to translational inhibition or increased protein degradation in seeds. Protein level (iBAQ, not log transformed) fold changes in CHX, MG132 and DMSO-treated samples were calculated relative to the baseline protein expression at 0h (control sample). The average protein level fold change for the subset of genes with either high or low PTR in seeds (see above) was subsequently determined for each time point and plotted separately for 'high' and 'low' PTR genes (Fig. 3f). To test for significant fold change differences between treatment time points, a one-way analysis of variance (ANOVA) and post-hoc Tukey HSD test was performed using the stats package in R (Extended Data Fig. 4i). For this, the data was normalized for changes in the proteome caused by seed germination, by calculating the $\log_2$ fold change in protein abundance (iBAQ) between seed samples treated with cycloheximide (CHX) or MG132 and the time-point matched DMSO control. Outlier values were removed for the boxplot visualization but are included in the ANOVA and post-hoc Tukey analysis.

***Co-expression analysis*:** To compare gene co-expression information in our dataset to prior knowledge about various associations between genes pairs, the Pearson correlations between all pair-wise gene combinations on both transcript and protein level (core dataset intersection; n = 5,043) was calculated. The correlation coefficient value for a given gene pair on transcript level was then plotted against the correlation coefficient value of the same pair on protein level (Extended Data Fig. 6a). Protein-protein association data was downloaded from STRING [30]

(www.string-db.org; 3702.protein.links.detailed.v10.5.txt.gz; download 07/2018), the experimental STRING subscores calculated for each gene pair and the information merged to the correlation value matrix. The scatterplot of transcript level versus protein level correlations was then divided into 50 x 50 bins along the x- and y-axis and the mean experimental STRING subscore for all gene pairs within a bin was calculated. Mean subscores were $\log_2$ transformed and visualized as a heatmap together with their marginal histograms (Extended Data Fig. 6a).

For a subset of genes, pair-wise co-expression was further investigated. Pairs used in this analysis were genes, that are either annotated as duplicated [31] or as protein interactors (AtPIN [33], download 08/2017), respectively. The Pearson correlation between the expression levels of gene-pairs across tissues was calculated for gene pairs with abundance measurements in more than ten matching tissues on both protein and transcript level (more than 10 pairwise complete observations). The duplicated gene set was further filtered for 3 unique peptide identification of each paralog. Top3 intensities were used as protein abundance measure for the duplicate gene set since the subsequent ratio comparison was also performed with top3 intensities. Paralog genes often have high sequence identity on amino acid level, which can lead to distorted ratios due to uneven assignment of shared (razor) peptides to only one of the paralogs. Duplicated genes (n = 3,612) were divided into three subsets: duplications caused by a whole genome duplication event (WGD, n = 2,104), local duplication (local, n = 408) or transposon-mediated duplication (transposed, n = 1,100). For each subset the distribution of Pearson correlation values between duplicates on either transcript or protein levels was plotted (Extended Data Fig. 6b). A set of random gene pairs (n = 27,353) was generated for comparison, filtered like the duplicated gene set (number of pairwise complete observation, 3 unique peptides) and the Pearson correlation value distribution displayed in the same way. To compare relative protein expression levels between paralogs, the ratio between the protein abundance (top3) of one paralog and the protein abundance (top3) of the other paralog was calculated (Extended Data Fig. 6c). As an example for the intensity proportion of paralog genes in different tissues, the top3 intensity (not log transformed) of MAC5A and MAC5B was plotted as proportion of the summed top3 intensity (not log transformed) of both paralogs (Extended Data Fig. 6c). For 57 paralog gene pairs with phenotypic information about asymmetric loss-of-function mutant combinations [32], we build the average top3 intensity proportion of the first paralog (dupl1/(dupl1+dupl2)) across 30 tissues and plotted them in descending order together with the phenotype information (Extended Data Fig. 6d).

A set of well-studied, stable protein complexes (26S proteasome, COP9 signalosome, Chaperonin, Cellulose synthase) was selected to estimate the protein level (iBAQ) co-expression expected for stable interaction partners. For this, the Pearson correlation coefficient values between all pair-wise subunit combinations of an individual complex were calculated and displayed in a density plot (Extended Data Fig. 6e). Based on this analysis, a Pearson correlation coefficient cutoff of >0.5 was subsequently used as indication for stable protein interactions. The protein interactor gene set (AtPIN) was used to test the recovery of annotated protein-protein interactions based on co-expression data from this study. Again, the distribution of Pearson correlation values between gene pairs (here interactors) on either transcript or protein level (iBAQ) was plotted. This was done for the whole AtPIN dataset (n = 57,152) and the following subsets: interactions identified by yeast-two hybrid experiments (Y2H, n = 7,621), by affinity-purification mass spectrometry (AP-MS, n = 17,982) or by both Y2H and AP-MS (n = 829).

***Size exclusion chromatography and complex analysis*:** Reproducibility between the three SEC experiments was tested by comparing the elution behavior of the same proteins. The coefficient of determination between protein peaks in the different experiments was above 0.7 (flower-root $R^2$ = 0.82; flower-leaf $R^2$ = 0.78; root-leaf $R^2$ = 0.7). Protein peak elution between the samples was also consistent across the whole SEC gradient range (Extended Data Fig. 7b).

To assign proteins to potential complexes, peak correlation profiling was performed. For this, the R package CCprofiler v0.1 was used to analyze the peptides.txt output table from MaxQuant [62,98]. Peptide table entries were restricted to the maximum molecular weight per gene name for the quantification of protein groups. This step removed peptides mapping to smaller isoforms of each gene. The restricted proteinGroups.txt served as the trace annotation input table for CCprofiler. The calibration table of the SEC column was based on the high molecular weight gel filtration calibration kit (GE Healthcare, Chicago, USA). The peptides table was split into the different tissues and the resulting tables converted into the input format for CCprofiler. All subsequent steps were performed separately for each tissue table.

The peptides table was imported, converted into a traces object, where a trace object is the SEC-elution profile of a specific peptide and annotated with the corresponding fractions and additional external information (e.g. molecular weight of the respective fraction, molecular weight of the respective protein) based on the aforementioned calibration and restricted proteinGroups tables. Traces not containing at least three consecutive peptide identifications across fractions, as well as traces with a sibling peptide correlation of less than 0.2 between

peptides originating from the same protein were filtered out. Based on the remaining traces, protein features (elution peak positional information) consisting of highly correlated peptides (correlation higher than 0.9) were detected using CCprofiler's sliding window algorithm. Peptides were randomly assigned to proteins in order to control the false discovery rate of protein features at 5% (random decoy model). Subsequently, protein traces were calculated by summing up the intensities of the top two most abundant peptide traces.

Next, complex hypotheses consisting of target and decoy protein complexes were constructed based on a previously-published mapping of Arabidopsis orthologues to mammalian protein complexes from CORUM [99,100]. Proteins were only mapped to the same decoy complex if they were not interacting with each other directly (minimum path length of two). These target and decoy complexes were used during the detection of complex features consisting of highly correlated proteins using the same sliding window algorithm as above in order to control the false discovery rate of complex features at 5% (target-decoy model).

Both protein features (elution peak positional information) and protein traces (quantitative information) were then used to generate an input matrix for weighted gene correlation network analysis (WGCNA) [101], which was used to find clusters of proteins with correlated SEC elution profiles. Protein traces were restricted to the ones, which were part of at least one protein feature. Each protein trace was duplicated for each protein feature it was part of and all intensities outside of the respective protein feature were set to zero. With this, protein traces were effectively split into separate elution peaks corresponding to distinct (sub)complexes. Afterwards, elution peaks, which met either of the following criteria were filtered out: (1) The absolute difference between the molecular weight at the apex of the elution peak and the monomer molecular weight of the corresponding protein was smaller than twice the monomer molecular weight. (2) The absolute difference between the molecular weight at the apex of the elution peak and the monomer molecular weight of the corresponding protein was less than 200 kDa. (3) The apex of the elution peak was in fraction 77 (void fraction). (4) The apex of the elution peak was in fraction 4 (monomer range).

The remaining elution peaks were restricted to the intersection with the core protein dataset across tissues described above (see section on tissue groups), generating the SEC WGCNA input dataset. At the same time, an additional WGCNA input dataset was generated based on the core protein dataset across tissues and restricted to the intersection with the SEC WGCNA input dataset. Proteins with several SEC elution peaks were duplicated to match the SEC WGCNA input dataset. WGCNA was carried out separately for each of these datasets. Signed

co-expression similarities were calculated between all pairs of proteins with at least seven pairwise-complete observations using the following formula:

$$s_{ij} = \frac{1 + cor(x_i, x_j)}{2}$$

Here, $s_{ij}$ is the signed co-expression similarity between two proteins $x_i$ and $x_j$ (based on their Pearson correlation across tissues). Adjacency matrices were calculated with $A = [s_{ij}]$ for several values of $\beta \in [1,30]$.

$$a_{ij} = s_{ij}^{\beta}$$

Here, $a_{ij}$ is the adjacency between two proteins $x_i$ and $x_j$.. The adjacency function parameter $\beta$ was selected to be 30 for the construction of signed protein co-expression networks. The topological overlap matrix $\Omega = [\omega_{ij}]$ of the two networks was calculated with:

$$\omega_{ij} = \frac{l_{ij} + a_{ij}}{min\{k_i, k_j\} + 1 - a_{ij}}$$

Here, $l_{ij} = \sum_u a_{iu} a_{uj}$ and $k_i = \sum_u a_{iu}$. Hierarchical clustering was performed on the topological-overlap-based dissimilarity matrices $D = [1 - \omega_{ij}]$ using average linkage. Correlation clusters ('Modules') were detected using adaptive branch pruning [102] using the 'Dynamic Hybrid' method set to respect the dendrogram topology during PAM operations and requiring at least 30 members per module. Similar clusters in the SEC or tissue dataset were merged, if the dissimilarity of their module eigengenes was smaller than 1-MPC or 1-MTC, respectively. MPC is the median peak correlation of the FDR-filtered complex features from CCprofiler (known complexes), while MTC is the median tissue correlation of the very same complex features (pairwise correlation between proteins mapping to a specific complex feature). The resulting modules for both datasets were then mapped to each other using a combination of Fisher's Exact test and manual curation. Enrichment of functional annotations from Corum [100], GO [73], KEGG [103] and Reactome [104] were calculated in each cluster using Fisher's Exact Test. All p-values were adjusted for multiple-testing by calculating FDRs [105] (Supplementary Data 8).

In order to quantify how well complexes can be detected using the tissue atlas (TA) or SEC WGCNA output, we calculated a summary statistic termed 'complex index'. Protein interactor information was downloaded from UniProt [6] (https://uniprot.org; 12/2018) and manually curated for the presence of large (> 4 subunits) and small (≤ 4 subunits) complexes (Extended Data Fig. 7e; Supplementary Data 9). The complex index (C) was calculated using the following formula:

$$C = \frac{Sm}{S} \times \frac{Sm}{M}$$

Here, $S_m$ is the number of subunits detected as present in the same module. *S* is the total number of subunits identified in the restricted datasets for either SEC or TA experiments and *M* is the total number of entries in the respective module. C is a measure of how well complex subunits co-occur (either by co-elution or by co-expression) and thus are detected in the same module and also quantifies the resolution of the detection method (module size). The complex index is 1, when all subunits of a complex are identified in the same module and no other proteins are contained in the module.

***Complex stoichiometry analysis*:** Absolute SEC protein elution traces of selected complexes (chaperonin, 26S Proteasome, CSN, CESA) were plotted using top3 intensities (not log transformed). Relative subunit proportions or stoichiometry for selected protein complexes within individual tissues of the TA dataset were calculated using top3 intensities (based on the 3 most abundant unique peptides). Top3 intensity rather than iBAQ intensity was used, in order to avoid distorted ratios caused by shared (razor) peptide assignment. In the case of paralog genes with high protein sequence similarity, peptides were not required to be unique for one gene, but rather for one 'paralog group'. The three most abundant peptides within a paralog group were then used for the top3 calculations.

For the example of the coatomer complex, the relative proportion of paralog genes in different tissues was calculated by plotting the top3 intensity (not log transformed, unique for gene) of each paralog as proportion of the summed top3 intensity (not log transformed) of all paralogs (Extended Data Fig. 7d). For the calculation of subunit stoichiometry ratios of selected complexes (chaperonin, 26S Proteasome, CSN, CESA), the top3 intensity (not log transformed) of a subunit was divided by the average top3 intensity (not log transformed) of all complex subunits. Subunit ratios were first calculated for each tissue and subsequently averaged across all 30 tissues (chaperonin, 26S Proteasome, CSN) or only across tissues where the respective subunits are mainly expressed (CESA4/7/8: node, internode, silique septum, silique valves; CESA1/3/6: all other) (Extended Data Fig. 7g).

***Subnetwork extraction:*** *De novo* network enrichment was performed to identify tissue- or tissue group-specific subnetworks and to link them to a molecular function. In contrast to gene set overrepresentation or gene set enrichment analysis [106], this approach is suited to identify previously uncharacterized functions from large-scale molecular interaction networks. KeyPathwayMiner [107] was used to extract tissue- and tissue group-specific subnetworks from

transcriptomics and proteomics datasets. To enrich subnetworks with proteins that deviate in their expression from other tissues, we employed z-scored expression values. These reflect how many standard deviations the expression in a given tissue is away from the expression range found in all other tissues:

$$z_{x,g} = \frac{x - \mu_{X\setminus x}}{\sigma_{X\setminus x}}$$

Here x is a specific tissue or tissue group of interest, g a gene $\mu$ the mean and $\sigma$ the standard deviation. Based on the following rule, a one-column indicator matrix I(x,g) was constructed as input for KeyPathwayMiner for each tissue or tissue group x:

$$I(x,g) = \begin{cases} 1 \; if \; |z_{x,g}| > 2 \\ 0 \; else \end{cases}$$

Here each row corresponds to a gene g. STRING {Szklarczyk, 2017 #513} *A. thaliana* network (v. 10.5, download 10/2018) was used as the molecular interaction network, considering only high confidence interactions with a score > 900. Subnetworks extracted via KeyPathwayMiner were made available for individual tissues or tissue groups as part of the ATHENA web application.

***Phosphorylation sites and motif analysis*:** Serine (S), threonin (T) and tyrosine (Y) content and p-site number of individual proteins was calculated based on the longest sequence in case of ambiguous isoform identifications (Fig. 5a, Extended Data Fig. 8g). Likewise, the schematic depiction of phosphorylation (p-site) localization in LATE EMBRYOGENESI ABUNDANT (LEA) and receptor-like kinase proteins was based on the p-site localization in the longest isoform sequence (Extended Data Fig. 8h,i). Functional domain assignment for receptor-like protein kinase (RLKs) sequences was done using smart (http://smart.embl-heidelberg.de/smart/batch.pl) [106] and p-sites were manually assigned to specific domains based on their localization in the protein (Extended Data Fig. 8i).

For motif analysis within our dataset, sequence windows of 15 amino acids centered on the identified class I p-sites (15 mers) were assigned to a motif class of either ‘proline-directed’, ‘acidic’, ‘basic’ or ‘other’ by following a binary decision tree [36] (Extended Data Fig. 8d). Assignment to these categories was done sequentially as follows: proline amino acid at position -1 (Proline-directed), 5 or more aspartic acid (D) or glutamic acid (E) at position +1 to +7 (acidic), arginine (R) or lysine (K) at position -3 (basic), D or E at position +1,+2 or +3 (acidic), 2 or more R or K at position -6 to -1 (basic) and otherwise (other).

P-sites with high confident localization scores (class I) were divided into S, T and Y p-site datasets. Motif extraction was performed separately for each motif class category (proline-directed, acidic, basic, other) using the motif-X algorithm, implemented in the rmotifx R package [35] (Supplementary Data 10). Cutoff settings were: min-seq.= 30, pval.cutoff = 1e-6 (S and T datasets); min-seq.= 10, pval.cutoff = 1e-5 (Y dataset). Motif-X calculates the fold-increase of a respective motif in the 'foreground' (p-site dataset) compared to the 'background' dataset (non-phosphorylated peptides). To avoid mass spectrometry-based residue biases, all peptides from our dataset that contained an STY amino acid and had only been identified in non-phosphorylated form were used as the background for all motif-X analyses (258,395 sequences). These peptide sequences were centered on STY and extended where necessary along the N- and C-terminal window to generate 15 mers. Position weight matrix sequence logos were drawn using the 'bits' method for amino acid sequences in the ggseqlogo R package [109] (Extended Data Fig. 8e, Supplementary Table 2). The motif-X fold increase for S motifs was plotted for motifs with 2, 3 and 4 fixed amino acid positions (Extended Data Fig. 8f).

Published p-site motifs were retrieved from PhosPhAT4.0 [9] and the Human Protein Reference Database [110] (HPRD; www.hprd.org) and divided into groups depending on the number of fixed amino acid positions within the motif sequence. These motifs were subsequently matched to the p-site or background dataset in order to calculate detection frequency and fold increase (p-site/background dataset) of a reported motif (Supplementary Data 10).

### *AGCVIII protein expression and mutant phenotypic* characterization

For the relative expression analysis of AGC kinases across tissues, the AGC1 and AGC3 subfamilies of Arabidopsis AGCVIII kinases were selected [21]. The summed total intensity of AGC1 and AGC3 subfamily kinases (top3, not log transformed) was calculated for each tissue. Relative protein amounts of individual kinases in different tissues were calculated as the top3 intensity (not log transformed, unique for gene) of each kinase in proportion to the summed top3 intensity (not log transformed) of all AGC1 and AGC3 kinases (Fig. 2c). For clear visualization only the D6PK family (D6PK, D6PKL1, D6PKL2, D6PKL3), AGC1.5, AGC1.6 and AGC1.7 were colored in the plot.

Mutant alleles of the *D6PK* family kinase genes, *d6pk-1*, *d6pkl1-1*, *d6pkl2-2* and *d6pkl3-2*, and their combinations were previously described [111]. Embryos were prepared omitting fixation with ethanol/acetic acid as previously described [112]. Embryos were analyzed with a Zeiss Axio Imager.M2, Axiocam 512 camera and 20x/0,8 Plan Apochromat objective using differential interference contrast (DIC).

For promoter:GUS constructs of *AGC1.5*, *AGC1.6* or *AGC1.7*, 2916 bp, 1207 bp, or 2214 bp fragments upstream from the respective start codon were cloned into the *Sal*I and *Eco*RI sites of pCAMBIA1391Z. Flowers from transgenic plants, obtained by the floral-dip method [50], were first incubated in 90% cold acetone for 15 min followed by a β-glucuronidase (GUS)-staining solution (50 mM sodium phosphate pH 7.0, 10 mM EDTA, 2 mM potassium ferricyanide, 2 mM potassium ferrocyanide, 0.1% Triton-X100, 0.5 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-glucuronide) over-night in the dark. Following a wash in 70% ethanol, flowers were incubated in ethanol/acetate (6:1) for removing of chlorophyll and then rehydrated in a graded ethanol series (90%, 70%, 50%, 30%). Samples were mounted in chloral hydrate/glycerol/water (8 g:1 ml: 2 ml) for imaging. The experiment was performed two times and GUS signal was observed for 12/18 *AGC1.5p::GUS*, 0/20 *AGC1.6p::GUS* and 12/18 *AGC1.7p::GUS* independent lines.

***ABA response of RCAR phosphomutants***

Preparation and analysis of Arabidopsis protoplasts was performed as described [113]. Briefly, protoplasts from three weeks-old Col-0 plants ($10^5$ protoplasts; 0.1 ml) were transfected with 5 µg of reporter construct (*pRD29B::LUC*), 3 µg of *p35S::GUS* plasmid as internal control and 3 µg of effector plasmid. The effector plasmids drive expression of respective RCAR and PP2C cDNAs under control of the 35S promoter [114]. RCAR10 phosphomimic reporter constructs ($RCAR10_{S32D}$; $RCAR10_{S113D}$) were obtained by site-directed mutagenesis using primer pairs S32D_F/S32D_R and S113D_F/S113D_R, respectively (Supplementary Data 11). The protoplast suspension was incubated at various levels of ABA as indicated and reporter expression was determined after 18 h of incubation at 25°C. Three biological replicates per data point were performed for each assay. All data points were normalized to the empty vector control. Structural modeling for RCAR10 was performed with SWISS-MODEL [115] using RCAR11 as template model (PDB ID 3k3k [45,116]) and modified with UCSF CHIMERA [117].

***Phenotypic characterization of QKY phosphomutants***

The pCAMBIA2300-based pQKY::mCherry:QKY construct was described previously [42]. The pQKY::mCherry:$QKY_{S262A}$ and pQKY::mCherry:$QKY_{S262E}$ plasmids were obtained using the Q5 site-directed mutagenesis kit (NEB, #E0554S) according to the manufacturers recommendation. Primers are Q5SDM S262A_F and Q5SDM S262A_R for pQKY::mCherry:$QKY_{S262A}$ and Q5SDM S262E_F and Q5SDM S262E_R for pQKY::mCherry:$QKY_{S262E}$ (Supplementary Data 11). All PCR-based constructs were confirmed by sequencing. Floral tissue for quantitative real-time PCR (qPCR) was harvested from plants grown under long day conditions. With minor changes, tissue collection, RNA extraction and qPCR analysis were performed as described

[118,119]. For detection of *mCherry:QKY*, expression by qPCR, primers mCherry-qRT-For/mCherry-qRT-Rev were employed (Table). Average *mCherry:QKY* expression was calculated relative to the expression of three control genes (AT4G33380; AT2G28390; AT5G46630) (Supplementary Data 11) and normalized to the wild type control for visualization. *A. thaliana* (L.) Heynh. var. Landsberg (*erecta*) (L*er*) was used as wild type strain. The likely null allele *qky*-9 (L*er*) has been described previously [43]. *qky*-9 mutant plants were transformed using Agrobacterium strain GV3101/pMP90 [120] and the floral dip method [50]. Transgenic T1 plants were selected on Kanamycin (50 µg/ml) and transferred to soil for further inspection. Plants were grown as described earlier [43]. Four out of 17 independent T1 transformants showed a wild type phenotype for the S262A transgenic line, while 13/17 T1 transformants displayed no (7/17) or partial (6/17) rescue, a notable decrease in functionality compared to the wild type construct (14/17 rescue, 2/17 partial rescue, 1/17 no rescue). For the S262E transgene, 11 out of 13 independent T1 transformants showed a wild type phenotype (1/13 partial rescue, 1/13 no rescue).

Floral organs were imaged using a Leica SAPO stereo microscope equipped with a digital MC 170 HD camera (Leica Microsystems GmbH, Wetzlar, Germany). Images were adjusted for color and contrast using ImageJ/Fiji software [121]. Confocal laser scanning microscopy of *qky*-9 *pQKY::mCHerry:QKY*, *qky*-9 *pQKY::mCHerry:QKY*$_{S262A}$ and *qky*-9 *pQKY::mCHerry:QKYS*$_{262E}$ six days-old seedling roots was performed with an Olympus FV1000 setup using an inverted IX81 stand and FluoView software (FV10-ASW version 01.04.00.09) (Olympus Europa GmbH, Hamburg, Germany) equipped with a water-corrected 40x objective (NA 0.9) at 3x digital zoom. Confocal high sensitivity detection (HSD) was employed involving two gallium arsenide phosphide photomultipliers (GaAsP PMTs) mounted equidistantly to the probe. The experiment was performed two times independently with 2 and 12 roots for *qky*-9 *pQKY::mCHerry:QKY*, 6 and 15 roots for *qky*-9 *pQKY::mCHerry:QKY*$_{S262A}$ and 15 roots for *qky*-9 *pQKY::mCHerry:QKYS*$_{262E}$.

### *Locus identifiers*

Gene locus identifiers are listed for all gene names mentioned in the manuscript (Supplementary Data 11).

### Online methods references


44 Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I. & Millar, A. H. SUBA: the Arabidopsis Subcellular Database. *Nucleic Acids Res* **35**, D213-218, doi:10.1093/nar/gkl863 (2007).

45 Nishimura, N. *et al.* Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science* **326**, 1373-1379, doi:10.1126/science.1181829 (2009).

46 Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in Arabidopsis. *Plant Cell* **2**, 755-767, doi:10.1105/tpc.2.8.755 (1990).
47 Johnson-Brousseau, S. A. & McCormick, S. A compendium of methods useful for characterizing Arabidopsis pollen mutants and gametophytically-expressed genes. *Plant J* **39**, 761-775, doi:10.1111/j.1365-313X.2004.02147.x (2004).
48 Sprunck, S. *et al.* Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* **338**, 1093-1097, doi:10.1126/science.1223944 (2012).
49 Karimi, M., Inze, D. & Depicker, A. GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci* **7**, 193-195 (2002).
50 Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**, 735-743 (1998).
51 Schmid, M. *et al.* A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37**, 501-506, doi:10.1038/ng1543 (2005).
52 Boyes, D. C. *et al.* Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* **13**, 1499-1510 (2001).
53 Bowman, J. L. *Arabidopsis : an atlas of morphology and development*. (Springer-Verlag, 1994).
54 Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254 (1976).
55 Ruprecht, B. *et al.* Optimized Enrichment of Phosphoproteomes by Fe-IMAC Column Chromatography. *Methods Mol Biol* **1550**, 47-60, doi:10.1007/978-1-4939-6747-6_5 (2017).
56 Marx, H. *et al.* A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* **31**, 557-564, doi:10.1038/nbt.2585 (2013).
57 Ruprecht, B., Zecha, J., Zolg, D. P. & Kuster, B. High pH Reversed-Phase Micro-Columns for Simple, Sensitive, and Efficient Fractionation of Proteome and (TMT labeled) Phosphoproteome Digests. *Methods Mol Biol* **1550**, 83-98, doi:10.1007/978-1-4939-6747-6_8 (2017).
58 Smith, P. K. *et al.* Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**, 76-85, doi:10.1016/0003-2697(85)90442-7 (1985).
59 Zolg, D. P. *et al.* PROCAL: A Set of 40 Peptide Standards for Retention Time Indexing, Column Performance Monitoring, and Collision Energy Calibration. *Proteomics* **17**, doi:10.1002/pmic.201700263 (2017).
60 Hahne, H. *et al.* DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* **10**, 989-991, doi:10.1038/nmeth.2610 (2013).
61 Bian, Y. *et al.* Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat Commun* **11**, 157, doi:10.1038/s41467-019-13973-x (2020).
62 Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319, doi:10.1038/nprot.2016.136 (2016).
63 Hanada, K. *et al.* sORF finder: a program package to identify small open reading frames with high coding potential. *Bioinformatics* **26**, 399-400, doi:10.1093/bioinformatics/btp688 (2010).
64 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652, doi:10.1038/nbt.1883 (2011).
65 Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**, 282-283, doi:10.1093/bioinformatics/17.3.282 (2001).
66 Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567, doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2 (1999).

67 Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat Protoc* **10**, 1567-1593, doi:10.1038/nprot.2015.101 (2015).
68 Toprak, U. H. *et al.* Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol Cell Proteomics* **13**, 2056-2071, doi:10.1074/mcp.O113.036475 (2014).
69 Onate-Sanchez, L. & Vicente-Carbajosa, J. DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC Res Notes* **1**, 93, doi:10.1186/1756-0500-1-93 (2008).
70 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).
71 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527, doi:10.1038/nbt.3519 (2016).
72 Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5**, 144-156, doi:10.1074/mcp.M500230-MCP200 (2006).
73 The Gene Ontology, C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**, D331-D338, doi:10.1093/nar/gkw1108 (2017).
74 Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13 Suppl 16**, S12, doi:10.1186/1471-2105-13-S16-S12 (2012).
75 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).
76 Olsen, J. V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648, doi:10.1016/j.cell.2006.09.026 (2006).
77 Uhlen, M. *et al.* Transcriptomics resources of human tissues and organs. *Mol Syst Biol* **12**, 862, doi:10.15252/msb.20155865 (2016).
78 Rijpkema, A. S., Vandenbussche, M., Koes, R., Heijmans, K. & Gerats, T. Variations on a theme: changes in the floral ABCs in angiosperms. *Semin Cell Dev Biol* **21**, 100-107, doi:10.1016/j.semcdb.2009.11.002 (2010).
79 Loytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol Biol* **1079**, 155-170, doi:10.1007/978-1-62703-646-7_10 (2014).
80 Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-552, doi:10.1093/oxfordjournals.molbev.a026334 (2000).
81 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).
82 van der Graaf, A. *et al.* Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc Natl Acad Sci U S A* **112**, 6676-6681, doi:10.1073/pnas.1424254112 (2015).
83 Gebert, D., Jehn, J. & Rosenkranz, D. Widespread selection for extremely high and low levels of secondary structure in coding sequences across all domains of life. *Open Biol* **9**, 190020, doi:10.1098/rsob.190020 (2019).
84 Camiolo, S., Melito, S. & Porceddu, A. New insights into the interplay between codon bias determinants in plants. *DNA Res* **22**, 461-470, doi:10.1093/dnares/dsv027 (2015).
85 Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* **102**, 14338-14343, doi:10.1073/pnas.0504070102 (2005).

86 Das, S. & Bansal, M. Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLoS One* **14**, e0212678, doi:10.1371/journal.pone.0212678 (2019).
87 Celaj, A. *et al.* Quantitative analysis of protein interaction network dynamics in yeast. *Mol Syst Biol* **13**, 934, doi:10.15252/msb.20177532 (2017).
88 Niederhuth, C. E. *et al.* Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17**, 194, doi:10.1186/s13059-016-1059-0 (2016).
89 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
90 Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-824, doi:10.1038/ng.2310 (2012).
91 Nakazawa, N. fmsb: Functions for Medical Statistics Book with some Demographic Data. R package version 0.6.3. *https://CRAN.R-project.org/package=fmsb* (2018).
92 R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing (https://www.R-project.org/.)* (2014).
93 Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society* **58**, 267-288 (1996).
94 Zhang, Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med* **4**, 136, doi:10.21037/atm.2016.03.35 (2016).
95 Knecht, W. Pilot Willingness to take Off Into Marginal Weather, Part II: Antecedent Overfitting With Forward Stepwise Logistic Regression. *Final Report, Federal Aviation Administration* (2005).
96 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
97 Groemping, U. Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software* **17**, 1-27, doi: 10.18637/jss.v017.i01 (2007).
98 Heusel, M. *et al.* Complex-centric proteome profiling by SEC-SWATH-MS. *Mol Syst Biol* **15**, e8438, doi:10.15252/msb.20188438 (2019).
99 McBride, Z., Chen, D., Reick, C., Xie, J. & Szymanski, D. B. Global Analysis of Membrane-associated Protein Oligomerization Using Protein Correlation Profiling. *Mol Cell Proteomics* **16**, 1972-1989, doi:10.1074/mcp.RA117.000276 (2017).
100 Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res* **38**, D497-501, doi:10.1093/nar/gkp914 (2010).
101 Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17, doi:10.2202/1544-6115.1128 (2005).
102 Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720, doi:10.1093/bioinformatics/btm563 (2008).
103 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092 (2017).
104 Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* **44**, D481-487, doi:10.1093/nar/gkv1351 (2016).
105 Hochberg, Y. B. a. Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289-300 (1995).
106 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

107 List, M. *et al.* KeyPathwayMinerWeb: online multi-omics network enrichment. *Nucleic Acids Res* **44**, W98-W104, doi:10.1093/nar/gkw373 (2016).

108 Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* **46**, D493-D496, doi:10.1093/nar/gkx922 (2018).

109 Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645-3647, doi:10.1093/bioinformatics/btx469 (2017).

110 Goel, R., Harsha, H. C., Pandey, A. & Prasad, T. S. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. *Mol Biosyst* **8**, 453-463, doi:10.1039/c1mb05340j (2012).

111 Zourelidou, M. *et al.* The polarly localized D6 PROTEIN KINASE is required for efficient auxin transport in Arabidopsis thaliana. *Development* **136**, 627-636, doi:10.1242/dev.028365 (2009).

112 Mayer, U. B., G.; Jurgens, G. Apical-basal pattern formation in the Arabidopsis embryo: studies on the role of the gnom gene. *Development* **177**, 149-162 (1993).

113 Moes, D., Himmelbach, A., Korte, A., Haberer, G. & Grill, E. Nuclear localization of the mutant protein phosphatase abi1 is required for insensitivity towards ABA responses in Arabidopsis. *Plant J* **54**, 806-819, doi:10.1111/j.1365-313X.2008.03454.x (2008).

114 Tischer, S. V. *et al.* Combinatorial interaction network of abscisic acid receptors and coreceptors from Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **114**, 10280-10285, doi:10.1073/pnas.1706593114 (2017).

115 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

116 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).

117 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

118 Box, M. S., Coustham, V., Dean, C. & Mylne, J. S. Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from Arabidopsis. *Plant Methods* **7**, 7, doi:10.1186/1746-4811-7-7 (2011).

119 Enugutti, B. *et al.* Regulation of planar growth by the Arabidopsis AGC protein kinase UNICORN. *Proc Natl Acad Sci U S A* **109**, 15060-15065, doi:10.1073/pnas.1205089109 (2012).

120 Koncz, C. & Schell, J. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Molecular and General Genetics MGG* **204**, 383-396, doi:10.1007/bf00331014 (1986).

121 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

122 Vizcaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**, D447-456, doi:10.1093/nar/gkv1145 (2016).

**Acknowledgements**

We thank the NGS@tum core facility for RNA sequencing, Rachele Tofanelli for help with imaging the ovules, Robert J. Schmitz for providing data access for the feature analysis and Maria Reinecke, Florian Bayer and Stefanie Galinec for MS measurements. This work was in part funded by the German Science Foundation (DFG, SFB924), a research fellowship to HS by the Japan Society for the Promotion of Sciences and a research fellowship to XC by the Chinese Research Council.

**Author contributions**

JM performed (phosho)proteomic and transcriptomic experiments under the supervision of BK. SR and HS performed AGC kinase experiments in plants under the supervision of GJ and CS. MP, AC and XC performed phosphomutant analysis under the supervision of EG and KS. PC and SS generated and provided plant material. JM, MF, MM, DL, SA, DPZ, TM, CD, AD and RRH performed data analysis under the supervision of BK, KFXM, PF, MB, TH and FJ. ML, PS and TS generated Arabidopsis resource databases under supervision of MW and JB. JM, CS and BK conceptualized the project and wrote the manuscript. All authors edited the manuscript.

**Competing interests**

MW and BK are founders and shareholders of OmicScouts GmbH and msAId GmbH. They have no operational role in the companies. MF and DPZ are founders and shareholders of msAId GmbH. TM and MB are employees and/or shareholders of Cellzome GmbH. The remaining authors declare no competing interests.

**Data and software availability**

The data supporting the findings of this study are available within the paper, the supplementary information and the public repositories. Source data for main Figure 1-5 and Extended Data Figure 1-9 are included. Transcriptome sequencing and quantification data are available at ArrayExpress (www.ebi.ac.uk/arrayexpress) under the identifier E-MTAB-7978. The raw mass spectrometric data and MaxQuant result files have been deposited to the ProteomeXchange Consortium via PRIDE [122], with the dataset identifier PXD013868.

**Corresponding author:** Bernhard Kuster (kuster@tum.de)

## Extended data figure legends

### Extended Data Figure 1 | Descriptive analysis of the multi-omic tissue atlas.

a, Pairwise global Pearson expression correlation analysis of all 30 tissues (n = 1 measurement per tissue) on the transcriptome level (lower triangle) and proteome level (upper triangle) using all identified gene loci. The data shows that proteins correlate more strongly between tissues than transcripts. Turquoise squares mark examples for morphologically highly similar tissues. Tissues are coloured as in Figure 1: Flower (light grey), seed (dark brown), pollen (yellow), stem (dark green), leaf (light green), root (dark grey), fruit (light brown), callus (magenta), cell culture (blue).

b, Scatter plots showing highly reproducible abundance measurements for transcript (upper panels) and protein (lower panels) in morphologically similar tissues that were marked in panel a, namely node (ND) versus internode (IND), leaf distal (LFD) versus leaf proximal (LFP) and root (RT) versus root upper zone (RTUZ). *R* denotes the Pearson correlation coefficient and n denotes the number of transcripts or proteins shown in the plots.

c, Percentage of genes encoded by a specific chromosome that were identified at the transcriptome, proteome or phosphoproteome level.

d, Percentage of Swiss-Prot and TrEMBL protein database entries as well as protein evidence categories from UniProt that were identified at the transcriptome, proteome or phosphoproteome level. Evidence level: (1) ‘protein evidence’, (2) ‘transcript evidence’, (3) ‘homology’, (4) ‘predicted’, (5) ‘uncertain’.

e, Comparison of protein identifications between an earlier Arabidopsis proteome study by Baerenfaller et al. based on 12 tissues [7], this study (30 tissues) and the number of protein-coding genes in Araport11.

f, iBAQ intensity distribution of proteins identified in this study. Proteins also identified by Baerenfaller et al. are projected into the same plot.

g, Left panel: Proportion of identified p-sites on S, T or Y residues with highly confident localization of the phosphorylation site within the identified peptide sequence (termed class I p-sites if the localization score is >0.75). Right panel: distribution of proteins for which phosphorylated S, T or Y residues were identified.

h, Left panel: Venn diagram comparing phosphoprotein datasets from van Wijk et al. [8], PhosPhAT4.0 and this study. Right panel: Venn diagram comparing p-site localization confidence between class I sites identified in this study and the low and high confidence datasets reported by van Wijk et al.

**Extended Data Figure 2 | Proteogenomics and dynamic range of transcript and protein expression.**

a, Number of identified N-terminal (NT) or C-terminal (CT) peptides of proteins in either unmodified or phosphorylated form.

b, Frequency of amino acids following the initiator methionine in N-terminal peptides with (-X) or without cleavage of the initiator methionine (M-X). X denotes the amino acid following the start codon.

c, Frequency of protein N-terminal acetylation for amino acids in (b). Because trypsin was used for protein digestion, the frequencies for arginine (R) and lysine (K) could not be determined (n.d.).

d, Distribution of peptide-based sequence coverage of proteins in individual tissues and for the combined dataset (tissue abbreviations as in Figure 1). Boxes contain 50% of the data and show the median as a black line. The upper and lower quartile ranges are shown as whiskers. The number of proteins is indicated for each tissue.

e, Pie charts showing the percentage of proteins identified by <3, 3-10 or >10 peptides either allowing shared (razor) peptides or restricting to unique peptides only.

f, Left panel: number of protein isoforms detected at the transcript and protein level compared to the number of all annotated isoforms in Araport11. Right panel: Number of multiple isoforms of the same gene distinguished at the peptide level.

g, Validation of protein isoform and short open reading frame (sORF) identification by comparing the tandem mass spectra from the tissue atlas to those of synthetic peptide reference standards. The normalized spectral contrast angle (SA) was used as a similarity metric (see methods). Candidate isoforms and sORFs were considered valid if the SA of the spectra was >0.7. This data is reported in Supplementary Data 3.

h, Amino acid sequence and mirror plots of tandem mass spectra for two peptides of the sORF BIP138_4. The spectra pointing upwards were collected from tissue digests, those pointing downwards were collected from synthetic peptides. The normalized spectral contrast angle (SA) and Pearson correlation coefficient (*R*) were used as similarity metrics (see methods) and indicate that both high scoring spectra (n=1 acquired spectra) are near identical, thus validating the identification of this sORF as a expressed protein.

i, Dynamic range of transcript expression (grey) and proportion of transcripts that were also identified at the protein level projected into this plot (blue). OM: orders of magnitude. Note that for lower abundance transcripts, fewer proteins were detected.

j, Dynamic range of protein expression and proportion of proteins with phosphorylation evidence. Note that protein expression spans 6 OM whereas transcript expression only spans 4 OM (panel i). Further note that phosphorylation was detected across the entire protein expression range.

k, Percentage of all annotated kinases (K), phosphatases (P), transcription factors (TF) and transcription regulators (TR) detected at the transcript, protein or phosphoprotein level. Numbers below the x-axis denote the number of genes for these protein classes in the AT genome.

**Extended Data Figure 3 | Descriptive analysis of transcript and protein expression in tissues.**

a. Distribution of expression specificity categories for protein and transcript identifications. See online methods for the definition of these categories. Briefly, there are very few transcripts and proteins that are only expressed in a single tissue. Note that the quantities of the shared transcripts or proteins can differ vastly between tissues (see below).

b, Left panel: protein identifications shared between flower (FL) and flower organs showing an almost complete qualitative overlap of proteins. Sepal (SP), petal (PT), stamen (ST), carpel (CP). Right panel: clustering of z-scored protein intensities showing distinct quantitative expression differences between flower organs.

c, Expression analysis of flower organ identity marker at the protein and transcript level. PISTILLATA (PI, green), APETALA3 (AP3, red), APETALA1 (AP1, orange), AGAMOUS (AG, blue). The expression of these markers is in line with the model of flower organ identity (AP1 expression marking sepal, AP1, AP3, PI marking petal, AG, AP3, PI marking stamen and AG marking carpel).

d, Total number of transcripts plotted against the total number of proteins detected in each individual tissue (n = 30 tissues) showing that the more genes are expressed as mRNAs, the more proteins can be detected in a tissue (Pearson correlation $R$ = 0.79). Tissues are coloured according to tissue groups as in Figure 1.

e, Cumulative abundance plots of intensity-ranked identifications of transcripts and proteins for five representative tissues. The five most abundant transcripts and proteins are listed in descending order for each tissue. Note that these are generally not the same. Further note that the characteristics of the plots are not the same for all tissues. In flower, the protein line rises more quickly than the transcript line. The opposite is true for pollen and a more even characteristic is observed in seed.

f, Distribution of shared and unique identifications among the 100 most abundant transcripts and proteins in each tissue. Note that relatively few proteins and transcripts are found together on

the list of the 100 most abundant transcripts and proteins. This demonstrates that the quantitative differences in transcript and protein expression are more important in defining a tissue than the qualitative expression of transcripts or proteins.

g, List of 11 proteins which were found as the most abundant protein (in at least one tissue) and their proportion of the total iBAQ intensity in each tissue. Note that individual proteins can represent up to 9% of the total protein in a given tissue.

h, Principal component analysis (PCA) of the core tissue proteomes and transcriptomes (i.e. the proteins and transcripts that were identified in every tissue) using z-scored abundances. Note that only about 30% of all protein and 20% of all mRNAs were detected in every of the 30 tissues despite the fact that all tissues were deeply profiled at both protein and transcript level. This shows that strong qualitative and quantitative expression differences exist between tissues. The PCA separates tissues into photosynthetically active versus inactive tissues (component 1) and separates pollen from all other tissues (component 2) implying that the molecular composition of pollen is particularly different from all other tissues.

i, Proportion of the total summed protein intensity for genes with specific subcellular compartment annotation (from SUBA [44], see methods) in the different tissue groups. The comparison of photosynthetically active and inactive tissues shows that e. g. a majority of the protein content in photosynthetically active tissues are contained in the plastids, whereas most protein is found in the cytosol for photosynthetically inactive tissues. Proteins with only one single subcellular compartment annotation were selected for the plot and the proportion of their iBAQ intensities were averaged for each tissue group. Nucleus (n = 1,393), endoplasmatic reticulum (n = 58), golgi (n = 68), peroxisome (n = 67), plastid (n = 525), mitochondrion (n = 317), vacuole (n = 71), cytosol (n = 385), cytoskeleton (n = 1), plasma membrane (n = 268), extracellular (n = 351).

**Extended Data Figure 4 | Relationships between transcript and protein levels.**

a, Pearson correlation (*R*) of transcriptome and proteome expression (core datasets; n = 5,043) for each tissue.

b, Pearson correlation (*R*) between measured and predicted protein abundance levels in all tissues. Predicted protein abundance levels were obtained from the best fitting feature selection model for each tissue (see methods). The number of genes used for the correlation analysis is indicated for each tissue.

c, Violin plots showing the spread in relative contribution of selected features to the prediction of gene-level protein abundance across tissues (n = 30 tissues) using our model. Violin shapes show the kernel density estimation of the data distribution and the median as white dot. Thick black bars denote the interquartile range.

d, Specific nucleotide sequence motifs in 5'UTRs of mRNAs contribute to the prediction of protein levels in a subset of tissues. Clustering tissues based on the presence or absence of detected 5'UTR motifs shows that several features are repeatedly selected for inclusion in the model while others appear to be more tissue-specific.

e, Based on the observation that the ratio of non-synonymous to synonymous nucleotide substitutions (Dn/Ds) between orthologous of *Arabidopsis thaliana* and *Arabidopsis lyrata* contributed to the prediction of protein levels (see above), we analysed this feature in more detail. Left panel: Distribution of the ratio of non-synonymous to synonymous nucleotide substitutions (Dn/Ds) for orthologous genes in *Arabidopsis thaliana* and *Arabidopsis lyrata*. The distribution is plotted for the example of 'leaf distal' (n = 6,447 genes). To compare evolutionarily conserved genes (defined by low non-synonymous to synonymous substitutions [Dn/Ds] ratios) and genes that evolve neutrally or are under positive selection (high Dn/Ds), we selected the bottom 5% and top 5% of the Dn/Ds ratio distribution, respectively. Right panel: evolutionarily conserved genes (low Dn/Ds ratio) show 10-20x higher protein abundance than genes under evolutionary pressure. Boxes contain 50% of the data and show the median as a black line. Whiskers denote 1.5 times the interquartile range. Outliers were omitted from the plot for clarity.

f, Time course analysis of median protein abundance changes upon cycloheximide (translation block, CHX) or MG132 (proteasome block) treatment vs time-matched DMSO control samples (see methods). Boxes contain 50% of the data and show medians as black lines. Whiskers denote 1.5 times the interquartile range. Outliers were omitted from the plot for clarity but were included in the statistical tests below. Shown are either all proteins in the experiment (n = 8,920, grey), proteins that have a high (n = 425, red) or low (n = 254, blue) PTR in seed (defined as in main Figure 3d). Differences between time points were tested for significance within each subset (all; high PTR; low PTR) using one-way ANOVA (analysis of variance) and the post-hoc Tukey HSD (Honestly significant difference) test. *** HSD p-Value < 0.001 (all_CHX8-CHX16: p-Value < 1e-7; all_CHX8-CHX24: p-Value < 1e-7; all_CHX16-CHX24: p-Value = 0.0002; highPTR_CHX8-CHX24: p-Value = 0.0003; lowPTR_CHX8-CHX16: p-Value = 0.0000004; lowPTR_CHX8-CHX24: p-Value < 1e-7; lowPTR_CHX16-CHX24: p-Value < 1e-7).

g, Representative images of seeds after 4 days of incubation with CHX, MG132 or DMSO control medium (n = 1). Germination was completely inhibited by CHX and partially inhibited by MG132 showing that the drug treatments were effective.

**Extended Data Figure 5 | Correlations between transcriptomes, proteomes and phosphoproteomes.**

a, Median protein to mRNA ratios (PTRs) across tissues plotted against the inter-tissue variation of these PTRs (expressed as median absolute deviation, MAD; proteins and transcripts had to be detected in at least 10 matching tissues to be included in the analysis).

Examples for genes with low transcript/high protein (rbcL and petA), and high transcript/low protein (IAA8 and IAA13) are marked by arrows. The bar blot shows the MAD range segmented into 5 quantiles each containing the same number of genes (coloured bars and dashed lines). It is evident that most genes have reasonably stable PTRs across tissues.

b, Same as panel a (dataset n = 14,069) but for transcript (left plot) and protein measurements (right plot). As can be seen, there is somewhat more variation in protein levels across tissues than there is mRNA variation (80% of all transcripts show a MAD of <1; 80% of all proteins show a MAD of 1.2). There is also more variation in the protein levels across tissues for low abundant proteins. This may in part due to technical limitations as low abundance proteins can generally be less accurately quantified.

c, Same as panel a but for the ratio of phosphorylation site versus protein abundance. P-sites and proteins had to be detected in at least 10 matching tissues to be included in the analysis (n = 13,793).

d, Same as panel b (dataset n = 13,793) but for phosphorylation site abundance. Note that p-site abundance shows somewhat greater variation across tissues than protein abundance (60% of all p-sites show MAD<1 compared to 80% of all proteins, see panel b). Again, this may in part due to technical limitations as p-site quantification is performed on a peptide level and does not benefit from aggregating multiple peptide quantifications into one value for protein quantification.

**Extended Data Figure 6 | Inferring redundant gene function and physical interactions from co-expression analysis.**

a, Scatter plot of Pearson correlation coefficients (*R*) as a measure for co-expression across tissues for all pairs of proteins (x-axis) and all pairs of transcripts (y-axis) (core dataset only, n = 5,043) along with their marginal histograms. Colours denote the $\log_2$-normalized STRING scores of individual gene pairs as a measure of known or predicted direct (physical) or indirect (functional) associations. The data show that strong co-expression of transcripts or proteins or both are more strongly related (physically or functionally) than transcripts and proteins that are not.

b, Co-expression analysis of duplicated genes (pairs had to be detected in at least 10 matching tissues to be included in the analysis). The density plots show the distribution of Pearson correlation coefficients (*R*) of co-expressed transcripts (grey) or proteins (blue) for genes that arose by whole genome duplications (WGD), local duplications (local) or transposon-mediated duplications (transposed). Randomly selected gene pairs are shown as control (random). Medians ($\tilde{x}$) are given and displayed as dotted lines. The data shows that there is substantial co-expression of duplicated genes implying that these genes likely have redundant functions.

c, Left panel: Protein level Pearson correlation coefficient (*R*) values (from panel b) for all duplicate gene pairs (WGD, local, transposed) plotted against the protein abundance ratio of each pair (average across 30 tissues, see methods). Blue arrows point out one example each for a high or low ratio of protein expression for the duplicated genes. Right panel: Example for tissue-resolved protein intensity proportions (top3, see methods) for the duplicate pair: MAC5A and MAC5B. The data shows that irrespective of the tissue, MAC5A is always much higher expressed than MAC5B. Tissues are coloured as in Figure 1.

d, Upper panel: ranked protein abundance ratio for selected duplicate pairs (average ± SD; n = 30) and annotated for phenotypic effects (lower panel) in the loss-of-function mutant for either duplicate 1 or duplicate 2 (+). Absence of a phenotypic effect is marked by (-). The data shows that asymmetric protein expression within duplicate pairs can be associated with the occurrence of a phenotype in the loss-of-function mutant of the higher expressed duplicate protein implying a dominant functional role of the more highly expressed protein. Blue arrows point out MAC5A/MAC5B and PHB3/PHB4 as examples.

e, Inference of physical protein-protein interactions from co-expression data. Distribution of pair-wise Pearson correlation coefficients (*R*) of co-expressed proteins across (at least 10) tissues that are subunits of selected protein complexes. An *R* value of >0.5 (shaded in grey) was chosen as a cut-off for the selection of proteins for subsequent analysis in order to make sure that proteins present in well characterized protein complexes are retained. CONSTITUTIVE PHOTOMORPHOGENESIS9 SIGNALOSOME (CSN), CELLULOSE SYNTHASE (CESA).

f, Recovery of annotated protein-protein interactions by co-expression analysis. Distribution of Pearson correlation coefficients (*R*) of pairs of transcripts (grey) or protein (blue) that are annotated to interact physically in the AtPIN database [33] (pairs had to be detected in at least 10 matching tissues to be included in the analysis). Subsets of the AtPIN database, namely interactions detected by the yeast two-hybrid (Y2H) method, by affinity purification - mass spectrometry (AP-MS) or by both (Y2H+AP-MS) are displayed separately. *R* values >0.5 are shaded in blue (protein). Median ($\tilde{x}$, dotted lines). The data shows that co-expression only recovers a minority of annotated physical interactions and that interactions supported by more than one line of experimental evidence also tend to show stronger co-expression.

**Extended Data Figure 7 | Inferring protein complexes and subunit stoichiometry from proteome correlation profiling using size-exclusion chromatography – mass spectrometry.**

a, Molecular weights of monomeric proteins (MW, determined from sequence) plotted against the MW determined from the apex of the elution profile for proteins identified by mass spectrometry in size-exclusion chromatography (SEC) fractions of flower tissue (sFL). The inset shows the MW calibration of the SEC column using a protein calibration standard (MW between

44 and 690 kDa). The MW distribution of proteins annotated in Araport11 is shown on top of the scatterplot. It is apparent, that a large number of proteins show a much higher apparent MW than what would be expected from their sequences (data points above the x=y line). This implies that these proteins engage in physical protein interactions that are sufficiently stable during SEC separation.

b, SEC traces of proteins from five well characterized protein complexes for flower, leaf and root tissue. Even though the resolution of SEC separations is not very high, it is apparent that the complex subunits show very strong co-elution behaviour and that the SEC separations of the 5 complexes are reproducible between tissues. Acetyl-CoA carboxylase n = 4 proteins; Cell division control protein 48 (CDC48) n = 3 proteins; Ribulose bisphosphate carboxylase oxygenase (RubisCO) n = 4 proteins; Prefoldin n = 6 proteins; Succinate-CoA synthetase (SCS) n = 3 proteins.

c, Intensity normalized SEC elution profile of proteins for flower tissue. Proteins are ordered based on the SEC fraction in which their intensity peaks and the data is displayed as heat map (n = 2,485 protein traces). Co-eluting proteins were grouped into so-called trace modules (see methods for details). Proteins in trace modules may represent members of protein complexes and thus serve as candidates for further experimental validation.

d, In order to quantify how well protein complexes can be detected using co-expression analysis from data in the tissue atlas (TA) or by SEC-MS a summary statistic termed ‘complex index’ was calculated (see methods). The complex index is 1, when all subunits of a complex are identified in the same module and no other proteins are contained in the module. The bar plots show examples for complex indices obtained from the different data sets and are divided into large (>4 subunits) and small (≤4 subunits) protein complexes (according to UniProt). The data shows that co-expression alone generates many candidates of interactors but that combining co-expression and SEC-MS analysis is an efficient way to prioritize candidates for follow-up experiments.

e, Subunit heterogeneity within the coatomer complex. The coatomer complex consists of 7 subunits, five of which (α, β, β’, ε and ζ) can be provided by 12 paralogs of these 5 genes. The plots show the protein proportions of these paralogs in all 30 tissues (data from tissue atlas). The data implies that the coatomer complex has a similar composition in most tissues. A notable exception are seed tissues in which subunit ζ-1 protein expression dominates over the two other paralogous proteins suggesting that the coatomer complex in seed tissue also preferentially contains the ζ-1 subunit. Tissues are coloured as in Figure 1.

f, Absolute SEC intensity traces of individual complex subunits for determining subunit stoichiometry. Examples from left to right: the chaperonin complex (flower, 8 proteins, ratio of all subunits: 1:1), the 26S proteasome core and lid (flower, 14+17 proteins, ratio of all subunits: 1:1), the COP9 Signalosome (flower, CSN; 8 proteins, ratio of all subunits: 1:1) and the

CESA1/3/6 complex (root, 3 proteins, ratio of all subunits: 1:1). Note that CSN3 and CSN5 were detected both as part of the CSN complex and in monomeric form.

g, Upper panels: total intensity of protein complex subunits across all tissues for the complexes shown in panel f (subunit intensities from the tissue atlas). Middle panels: Relative proportion of the complex subunits within each tissue. Lower panels: Estimation of subunit stoichiometry using the average ± SD (n = 30 tissues) of the proportions of subunits across tissues (see methods). For the CESA complex, the ratios were calculated for the subunit combinations CESA1/3/6 and CESA4/7/8. The data shows that the stoichiometries determined from the tissue expression data are generally well aligned with the expected 1:1 ratio of subunits in these complexes. As noted above, a substantial amount of CSN5 was detected as a monomer in the SEC analysis and the tissue expression atlas also shows higher relative expression of this protein compared to all other complex partners. This suggests that this protein is produced in excess over what is required for the COP9 complex (as has been observed by others before) and may therefore imply an additional function within the cell.

**Extended Data Figure 8 | Kinases, phosphatases and phosphorylation motifs.**

a, Percentage of annotated kinases and phosphatases family members detected at the protein or phosphoprotein level. Brackets denote the number of genes in each family in the Arabidopsis genome

b, Tissue-resolved combined intensity (i.e. protein expression) of families of kinases (left) and phosphatases (right). Tissues are coloured as in Figure 1. Note that several tissues (notably pollen) stand out in terms of the expression of kinases and phosphatases implying that these tissues are particularly active in phosphorylation-mediated dynamic signalling.

c, Upper panel: Pie chart of protein expression specificity categories (see methods for definition) for kinases and phosphatases. Lower panels: Distribution of tissue-enhanced kinases and phosphatases across the 30 tissues showing that several tissues (notably pollen) stand out in terms of the expression of certain kinases and phosphatases implying tissue-specific signalling.

d, Pie charts showing the proportion of ‘proline-directed’, ‘acidic’, ‘basic’ and ‘other’ motif categories for phosphorylated serine (pS), threonine (pT) and tyrosine (pY) residues. Only class I p-sites (localization score > 0.75, methods) were considered in this analysis.

e, Example motif logo plots for ‘proline-directed’, ‘acidic’, ‘basic’ and ‘other’ motifs. P-site motifs were identified using the motif-X algorithm (see Supplementary Table 2 for all 266 motifs). n denotes the number of phosphorylation sites that contain the respective motif; fc denotes the fold change (i. e. enrichment) of the motif in phosphorylated vs unmodified peptides (see methods).

f, Enrichment of 'proline-directed' (yellow), 'acidic' (red), 'basic' (blue) and 'other' (grey) sequence motifs (circles) in the serine p-site dataset versus the same motifs detected in the background dataset of unmodified peptides (see methods). Motifs are shown for 2, 3 and 4 fixed amino acid positions. The p-site in each motif example is underlined. X denotes any amino acid.

g, Number of identified p-sites for a given protein plotted against the sequence lengths of the same protein. LATE EMBRYOGENESIS ABUNDANT (LEA) proteins are marked by a circle and arrow.

h, Schematic representation of the LEA protein sequences (black bars). Pink marks denote phosphorylated and blue marks unphosphorylated STY residues. It is evident, that almost all STY residues in LEA proteins can be phosphorylated.

i, Schematic representation the sequences and domain topology of the receptor-like kinases STRUBBELIG RECEPTOR FAMILY4 (SRF4), FERONIA (FER) and CHITIN RECEPTOR KINASE1 (CERK1). It is apparent that p-sites often preferentially occur in specific domains, notably the juxtamembrane domain. Protein sequence regions covered by identified peptides are marked in blue and p-sites are marked in pink.

**Extended Data Figure 9 | Functional analysis of phosphorylation mutants of RCAR10 and QKY.**

a, Phosphorylation site (p-site) localization within the structure of REGULATORY COMPONENTS OF ABA RECEPTOR10 (RCAR10). The RCAR10 structure (blue) was modelled using the RCAR11 protein crystal structure (cornflowerblue) as a template [45]. ABA-binding loops are shown in turquoise, p-sites in pink and ABA ligand in yellow.

b, RCAR10 expression across tissues at the protein (blue, iBAQ), transcript (grey, TPM) and p-site (pink, intensity) level.

c, Tissue-resolved total protein intensity and relative proportions of the members of the protein phosphatase 2C (PP2C) co-receptor family. Note that seed tissues stand out in terms of overall expression as well as the dominance of AHG1 in these tissues.

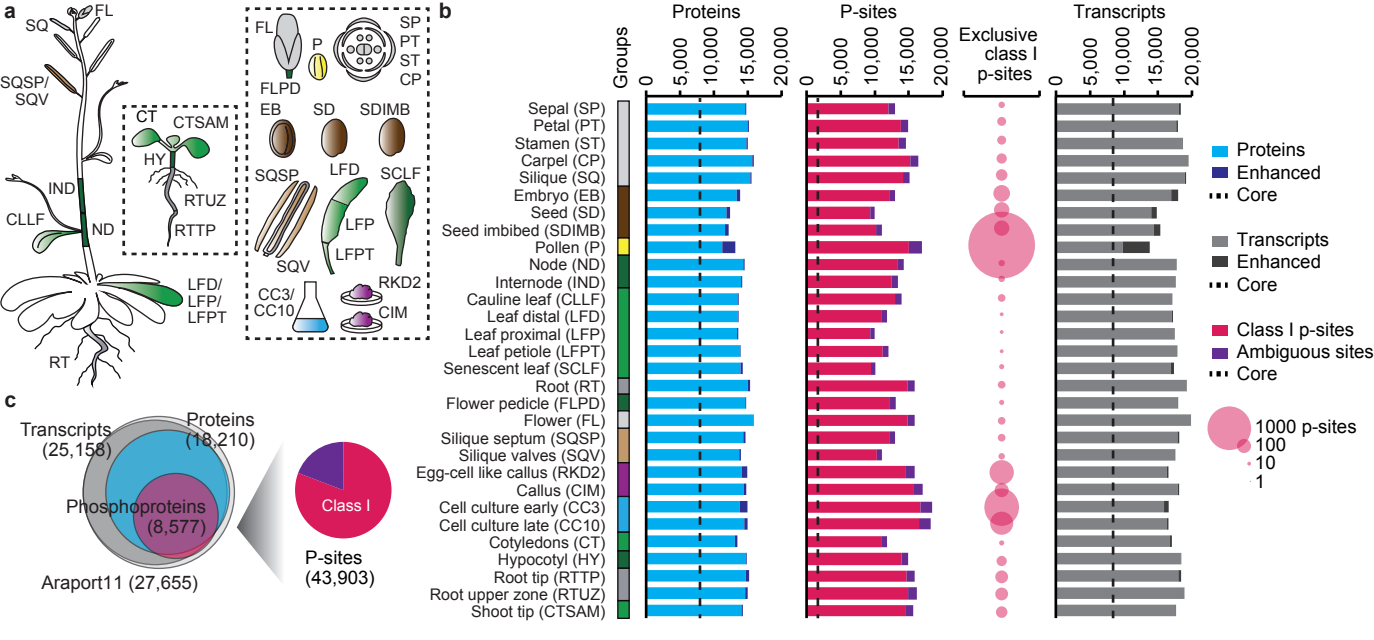
d, Measurement of ABA response when expressing RCAR or RCAR10 phosphomimetic mutant variants in combination with different PP2C co-receptors in protoplasts (see methods). Columns display the average ABA response (±SD, n = 3) and grey dots indicate individual measurements. The data shows that co-expression of the phosphatases HAI1-3 lead to similar responses of the phophomimic mutants, whereas other co-expressed phosphatases show diverse responses.

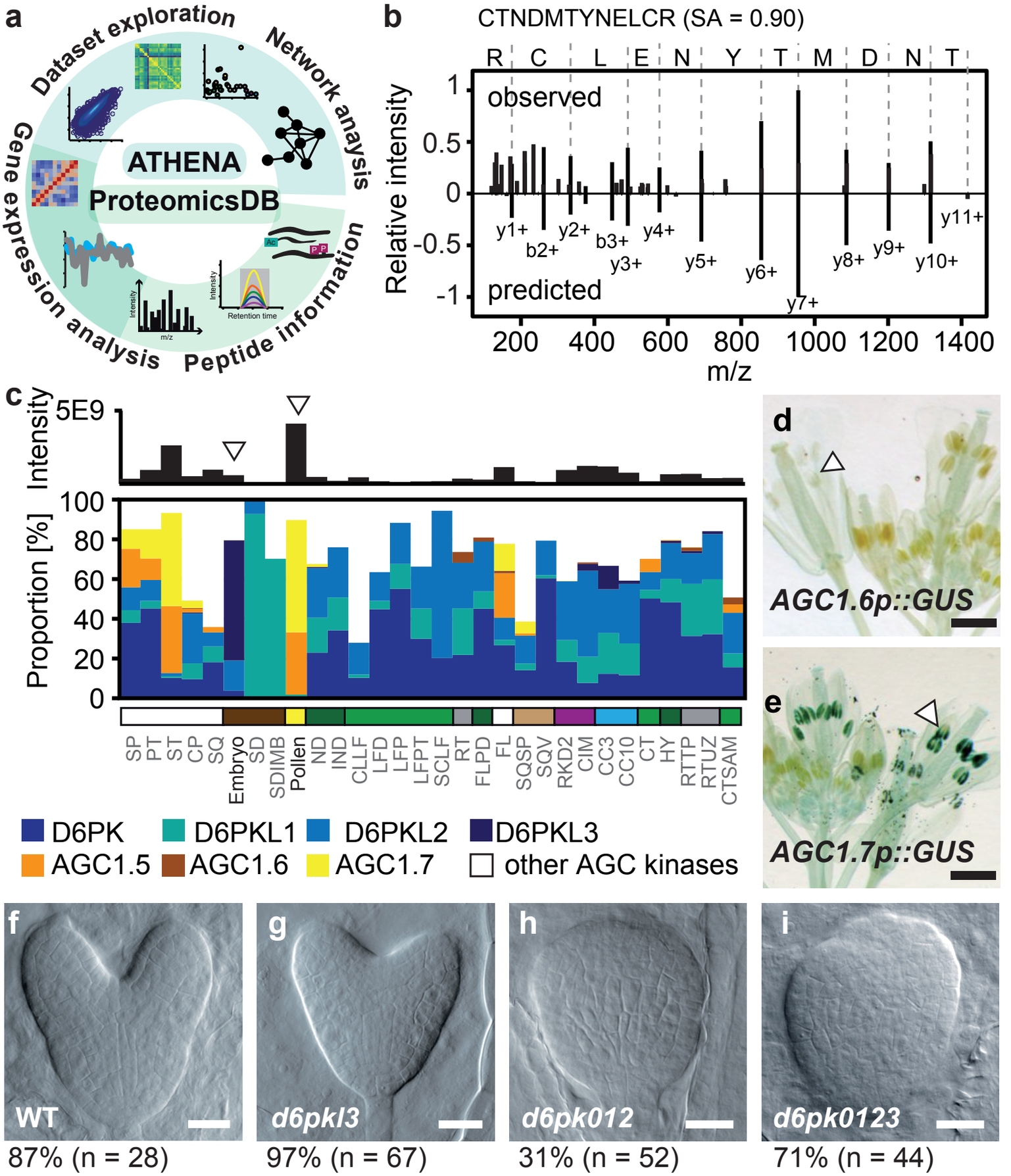
e, QUIRKY (QKY) expression across tissues at the protein (blue, iBAQ), transcript (grey, TPM) and p-site (pink, intensity) level.

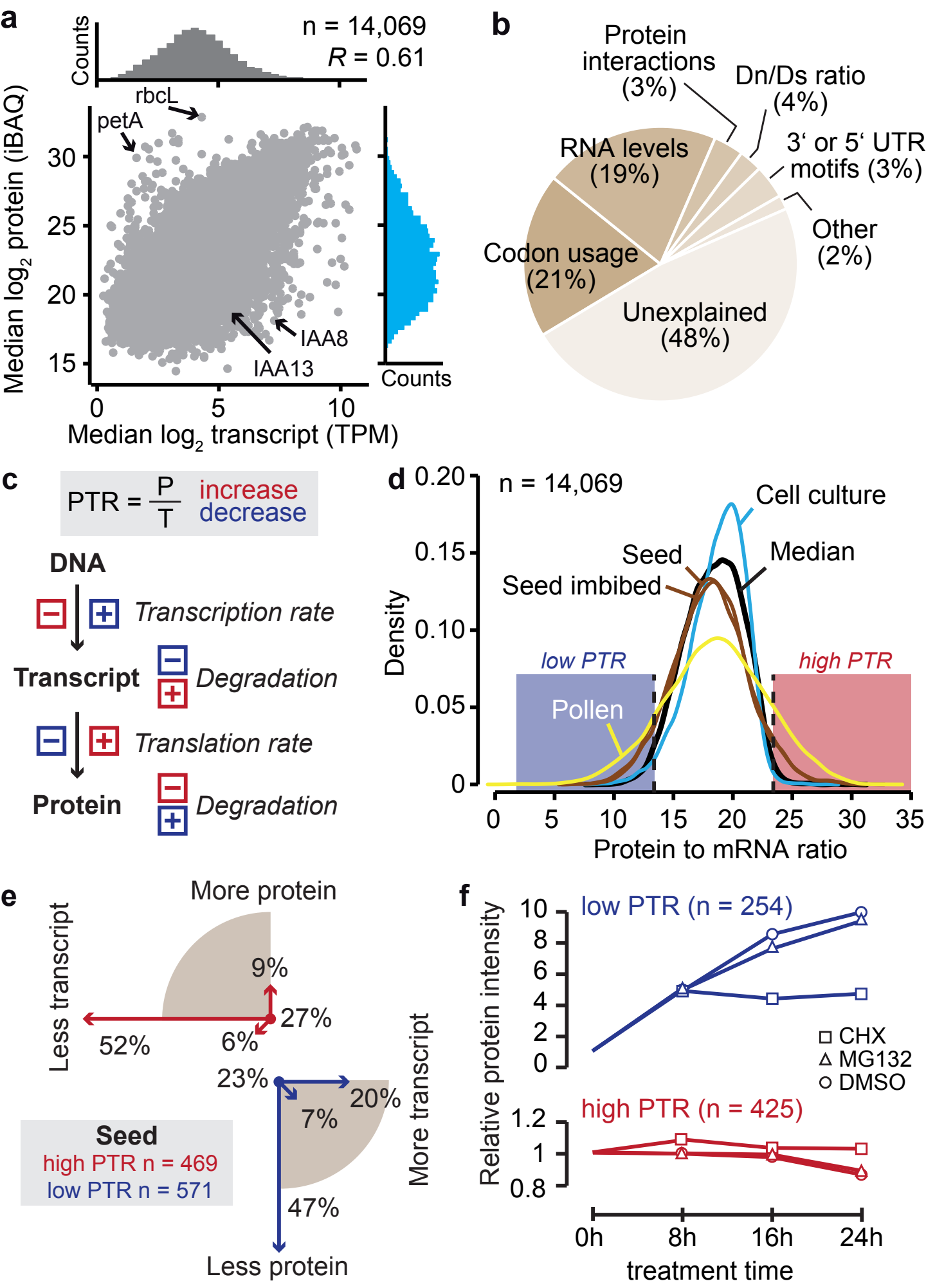
f, Members of the MULTIPLE C2 AND TRANSMEMBRANE DOMAIN-CONTAINING PROTEIN (MCTP) family clustered by sequence similarity (left) and schematic representation of their domain structures along with detected p-sites (right). MCTP11a, 12 and 13 were not detected in this study (n.d.). QKY (MCTP15) is marked in bold.

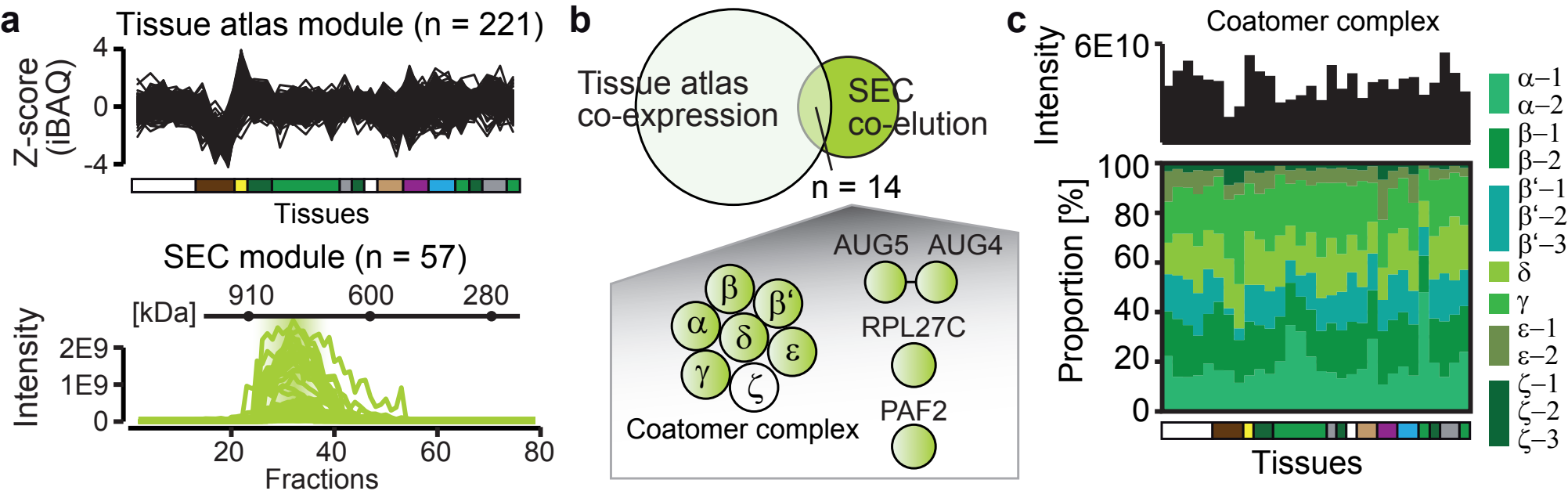
g, Number of independent transgenic plant lines (*qky*-9 mutant background) transformed with WT (QKY), phosphomutant (SA) or phosphomimic (SE) constructs that show complete, partial or no rescue of the mutant phenotype. qPCR results (average ± SD; individual data points as grey dots) show the relative transgene expression in wild type (WT), *qky*-9 mutant and selected transgenic lines. QKY wild type (grey), $QKY_{S262A}$ (SA; blue), $QKY_{S262E}$ (SE, purple).

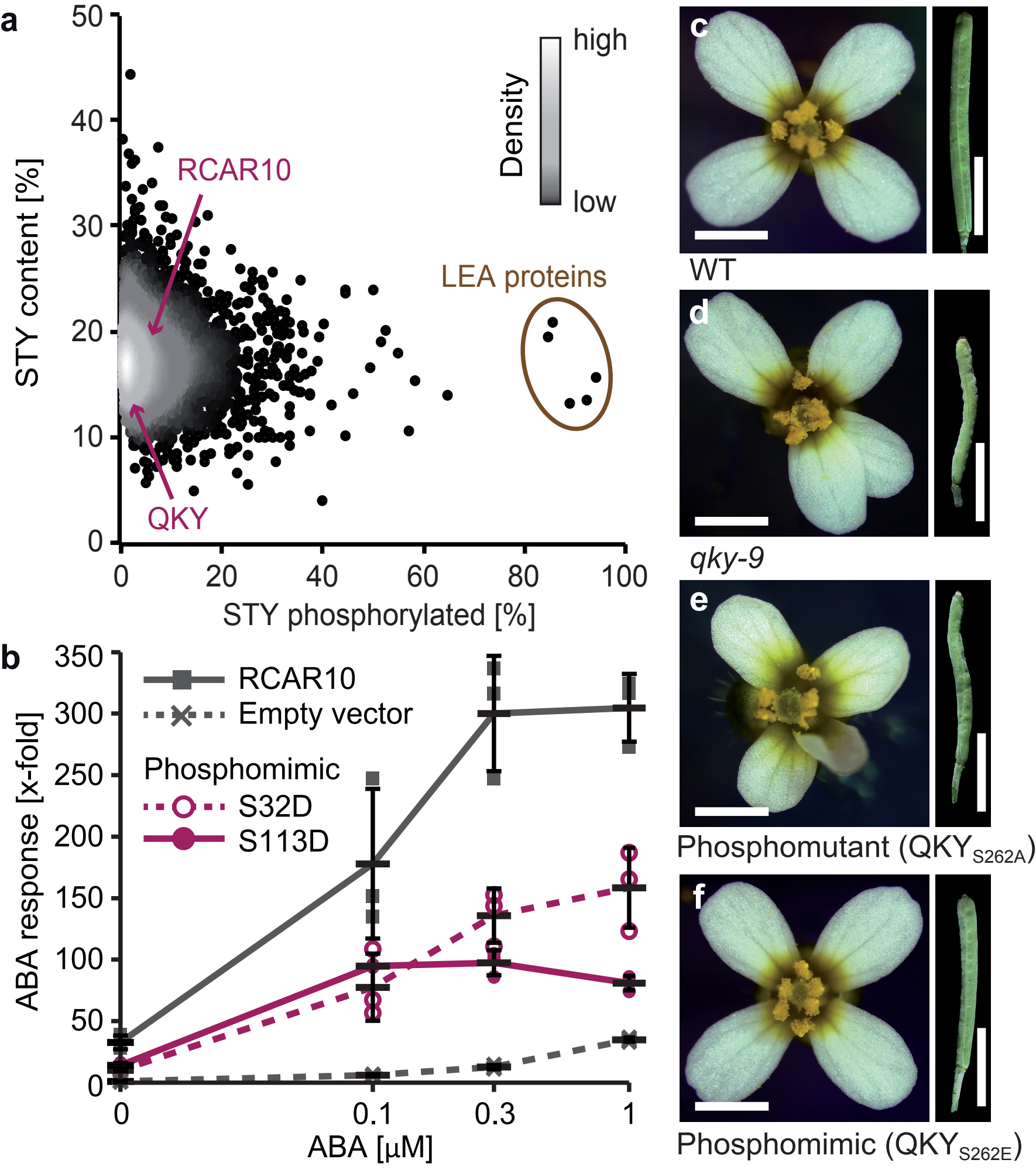
h-j, Representative confocal images of six days-old *qky*-9 *pQKY::mCherry:QKY* (WT QKY; n = 14 roots), *qky*-9 *pQKY::mCherry:QKYS262A* (phosphomutant; n = 21 roots), *qky*-9 *pQKY::mCherry:QKYS262E* (phosphomimic; n = 15 roots) root epidermal cells of the meristematic zone. The punctate signal along the cell circumference shows the expected localization of QKY protein. Arrows indicate punctate structures. Scale bars 5 µm.


a
FL
SQ
SQSP/
SQV
IND
CLLF
ND
CT
CTSAM
HY
RTUZ
RTTP
LFD/
LFP/
LFPT
RT
FL
P
SP
PT
ST
CP
FLPD
EB
SD
SDIMB
SQSP
LFD
SCLF
LFP
LFPT
SQV
RKD2
CC3/
CC10
CIM
b
Groups
Proteins
P-sites
Exclusive
class I
p-sites
Transcripts
0
5,000
10,000
15,000
20,000
Sepal (SP)
Petal (PT)
Stamen (ST)
Carpel (CP)
Silique (SQ)
Embryo (EB)
Seed (SD)
Seed imbibed (SDIMB)
Pollen (P)
Node (ND)
Internode (IND)
Cauline leaf (CLLF)
Leaf distal (LFD)
Leaf proximal (LFP)
Leaf petiole (LFPT)
Senescent leaf (SCLF)
Root (RT)
Flower pedicle (FLPD)
Flower (FL)
Silique septum (SQSP)
Silique valves (SQV)
Egg-cell like callus (RKD2)
Callus (CIM)
Cell culture early (CC3)
Cell culture late (CC10)
Cotyledons (CT)
Hypocotyl (HY)
Root tip (RTTP)
Root upper zone (RTUZ)
Shoot tip (CTSAM)
Proteins
Enhanced
Core
Transcripts
Enhanced
Core
Class I p-sites
Ambiguous sites
Core
1000 p-sites
100
10
1
c
Transcripts
(25,158)
Proteins
(18,210)
Phosphoproteins
(8,577)
Araport11 (27,655)
Class I
P-sites
(43,903)


a
Dataset exploration
Network analysis
Peptide information
Gene expression analysis
ATHENA
ProteomicsDB
Intensity
m/z
Retention time
b
CTNDMTYNELCR (SA = 0.90)
R C L E N Y T M D N T
observed
predicted
Relative intensity
y1+
b2+
y2+
b3+
y3+
y4+
y5+
y6+
y7+
y8+
y9+
y10+
y11+
m/z
c
Intensity
5E9
Proportion [%]
SP PT ST CP SQ Embryo SD SDIMB Pollen ND IND CLLF LFD LFP LFPT SCLF RT FLPD FL SQSP SQV RKD2 CIM CC3 CC10 CT HY RTTP RTUZ CTSAM
D6PK
D6PKL1
D6PKL2
D6PKL3
AGC1.5
AGC1.6
AGC1.7
other AGC kinases
d
AGC1.6p::GUS
e
AGC1.7p::GUS
f
WT
87% (n = 28)
g
d6pkl3
97% (n = 67)
h
d6pk012
31% (n = 52)
i
d6pk0123
71% (n = 44)


a
n = 14,069
R = 0.61
Counts
rbcL
petA
IAA8
IAA13
Median log2 protein (iBAQ)
Median log2 transcript (TPM)
b
Protein interactions (3%)
Dn/Ds ratio (4%)
3‘ or 5‘ UTR motifs (3%)
Other (2%)
RNA levels (19%)
Codon usage (21%)
Unexplained (48%)
c
PTR = P/T
increase
decrease
DNA
Transcription rate
Transcript
Degradation
Translation rate
Protein
Degradation
d
n = 14,069
Cell culture
Seed
Median
Seed imbibed
low PTR
high PTR
Pollen
Density
Protein to mRNA ratio
e
More protein
Less transcript
More transcript
Less protein
9%
27%
52%
6%
23%
20%
7%
47%
Seed
high PTR n = 469
low PTR n = 571
f
low PTR (n = 254)
high PTR (n = 425)
CHX
MG132
DMSO
Relative protein intensity
0h
8h
16h
24h
treatment time


a
Tissue atlas module (n = 221)
Z-score (iBAQ)
4
0
-4
Tissues
SEC module (n = 57)
[kDa]
910
600
280
Intensity
2E9
1E9
0
20
40
60
80
Fractions
b
Tissue atlas co-expression
SEC co-elution
n = 14
AUG5
AUG4
RPL27C
PAF2
α
β
β‘
γ
δ
ε
ζ
Coatomer complex
c
Coatomer complex
Intensity
6E10
Proportion [%]
100
80
60
40
20
0
Tissues
α–1
α–2
β–1
β–2
β‘–1
β‘–2
β‘–3
δ
γ
ε–1
ε–2
ζ–1
ζ–2
ζ–3


a
STY content [%]
STY phosphorylated [%]
RCAR10
QKY
LEA proteins
Density
high
low
b
ABA response [x-fold]
ABA [µM]
RCAR10
Empty vector
Phosphomimic
S32D
S113D
c
WT
d
qky-9
e
Phosphomutant (QKY S262A)
f
Phosphomimic (QKY S262E)