

RESEARCH

Open Access

Mapping gene activity of *Arabidopsis* root hairs

Ping Lan^{1†}, Wenfeng Li^{1†}, Wen-Dar Lin¹, Simonetta Santi² and Wolfgang Schmidt^{1,3,4*}

Abstract

Background: Quantitative information on gene activity at single cell-type resolution is essential for the understanding of how cells work and interact. Root hairs, or trichoblasts, tubular-shaped outgrowths of specialized cells in the epidermis, represent an ideal model for cell fate acquisition and differentiation in plants.

Results: Here, we provide an atlas of gene and protein expression in *Arabidopsis* root hair cells, generated by paired-end RNA sequencing and LC/MS-MS analysis of protoplasts from plants containing a *pEXP7-GFP* reporter construct. In total, transcripts of 23,034 genes were detected in root hairs. High-resolution proteome analysis led to the reliable identification of 2,447 proteins, 129 of which were differentially expressed between root hairs and non-root hair tissue. Dissection of pre-mRNA splicing patterns showed that all types of alternative splicing were cell type-dependent, and less complex in *EXP7*-expressing cells when compared to non-root hair cells. Intron retention was repressed in several transcripts functionally related to root hair morphogenesis, indicative of a cell type-specific control of gene expression by alternative splicing of pre-mRNA. Concordance between mRNA and protein expression was generally high, but in many cases mRNA expression was not predictive for protein abundance.

Conclusions: The integrated analysis shows that gene activity in root hairs is dictated by orchestrated, multilayered regulatory mechanisms that allow for a cell type-specific composition of functional components.

Background

Systems-wide exploration of 'omics' data obtained at different molecular levels provides a way to understand physiological or developmental processes. The fidelity of large-scale analysis of gene activity has dramatically increased because of new technologies in transcriptional profiling such as RNA sequencing (RNA-seq) and advances in mass spectrometry (MS) techniques for protein profiling, allowing more accurate detection of expressed genes. In multicellular organisms, the correct assembly of disparate datasets derived from parallel profiling experiments is often obscured by an amalgam of different tissues or cell types, compromising the comparability of these data. Despite the technical improvements in high-throughput assays, genome-wide exploration of gene activity at the resolution of single cell types is still a challenging task.

Root hairs, which differentiate from specialized cells in the epidermis, represent a well-explored model for cell differentiation and growth. Root hairs are crucial for the

uptake of water and nutrients, and are important in micro-organism/root interactions, thereby contributing to plant fitness. In *Arabidopsis*, root hairs are organized in cell files in a position-dependent manner. The fate of cells is dependent on their contact; cells that span the cleft of two underlying cortical cells (H position) develop into hair cells, whereas cells that are in contact with only one cortical cell (N position) develop into a non-hair cell [1]. Cell fate is determined by a complex mechanism that includes the reciprocal cell-to-cell movement of transcription factors, initiated by a positional signal that is presumably stronger in the H position and represses the expression of *WEREWOLF (WER)* in the future hair cells [2-4]. Root-hair formation commences with the formation of a bulge at the basal end of the epidermal cell, followed by highly polarized tip growth that results in rapid elongation of the hair. The formation of root hairs necessitates the concerted action of numerous players controlling an array of processes including reorganization of the cytoskeleton, which is guided by ROP-GTP signaling, auxin distribution, vesicle trafficking, cell wall reassembly, production of reactive oxygen species, and the establishment of ion gradients to allow proper growth of the cell [5,6]. By comparing the transcriptional profiles of the tip growth-defective mutant

* Correspondence: wosh@gate.sinica.edu.tw

† Contributed equally

¹Institute of Plant and Microbial Biology, Academia Sinica, 11529 Taipei, Taiwan

Full list of author information is available at the end of the article

rhd2 with those of the wild type, a suite of 606 genes with putative functions in root-hair morphogenesis was previously identified, yielding the first genome-wide overview of root-hair differentiation at the transcriptional level [7]. A cell type-specific gene-expression profiling study was conducted by Birnbaum *et al.* [8], using fluorescence-activated cell sorting (FACS) of plant-root protoplasts. In that study, 10,492 genes were detected in the root, and mapped to five different tissues in three developmental root zones. This analysis was later extended into a spatio-temporal expression atlas of *Arabidopsis* roots, investigating 14 non-overlapping cell types and 13 root sections representing different developmental stages. The results of that study identified complex and partly fluctuating transcriptional patterns that determine cell-identity programs [9]. Cell type-specific expression profiling in response to environmental conditions identified coordinated responses in distinct cell types and showed that this approach dramatically increases the detection sensitivity for transcriptional changes compared with studies using whole roots as experimental material [10]. It was further shown that cell type-specific transcription is largely dependent on environmental conditions, with the epidermis showing the least conserved gene expression when the transcriptional profile of stressed plants was compared with that of plants grown under standard conditions [11].

A gene regulatory network, identified by comprehensive transcriptional profiling of epidermal cells from several cell-fate mutants, provided detailed insights into the transcriptional machinery regulating cell specification and root-hair development at systems level [12]. By integrating genetic, genomic, and computational analyses, this approach yielded a subset of 208 'core' root-hair genes from a large set of genes that were differentially expressed between root-hair and non-root-hair cells. These genes could be organized within the network by using information that was obtained from transcriptional analysis of root epidermis mutants after the system was perturbed by hormone treatment. A Bayesian modeling approach in combination with transcriptional profiles from different developmental stages further positioned the genes within the network.

Although transcriptome analysis has yielded biological meaningful information that awaits full exploitation, several studies have shown that transcripts do not reliably predict the abundance of proteins, indicating the need for a multifaceted, integrative analysis that includes dissection of both the transcriptome and proteome to obtain a holistic picture of gene activity. Protein profiling of root protoplasts identified only modest correlation of protein and mRNA abundance [13], indicating complex regulation of gene expression. To further extend the knowledge of root-hair differentiation, we conducted an in-depth analysis at

several levels of gene expression without introducing either conceptual (that is, by experiments being carried out in different laboratories) or technical (that is, by amplification of RNA samples or by fixed probe sets) bias. This dataset includes quantitative information on transcript abundance and composition, cell type-dependent splicing variants, proteins that differentially accumulate in root hairs, and post-translational modifications. The easily mined dataset provides a guide towards understanding how gene function is integrated to control the development and function of root hairs.

Results

Quantification of gene expression in *Arabidopsis* root hairs

Expression of the gene encoding for expansin 7 (*EXP7*) is tightly linked to root-hair initiation and elongation [14]. In five-day-old seedlings carrying a construct containing the *EXP7* promoter sequence positioned upstream of a green fluorescent protein (GFP) reporter, expression of the *pEXP7-GFP* chimera was observable in developing and mature root hairs, but no expression was detected before the initiation of bulge formation (Figure 1). Expression of *EXP7* was restricted to trichoblasts (cells that will develop a root hair), and no GFP signal was seen in non-hair cells (Figure 1a,c). Roots from plants containing the *pEXP7-GFP* construct were treated with cell wall-digesting enzymes, and GFP-expressing protoplasts (referred to as EXP7 cells) were separated from non-GFP protoplasts (comprising all root tissues except root hair cells) using a FACS system equipped with a cooling mechanism. The percentage of EXP7-expressing protoplasts was about 1% in a typical sorting experiment (see Additional file 1, Figure S1). To avoid bias associated with mandatory RNA amplification in downstream applications, RNA was extracted from a pooled population of protoplasts derived from several sorting experiments.

The transcriptome of EXP7 and non-GFP cells was analyzed using RNA-seq on a genome analyzer platform. In total, 126.6 (EXP7) and 155.4 (non-GFP) million paired-end reads from two independent experiments were mapped to gene models annotated in TAIR10 with at least 95% identity, corresponding to 20,822 and 21,358 transcripts for EXP7 and non-GFP cells, respectively, with a read number of five or more in each of the runs (Figure 2; see Additional file 2, Data set S1). The experiments showed remarkably high similarity both between the two cell types and between the two biological repeats for each cell type. A core of 19,929 transcripts from a total of 22,251 was identified in both EXP7 and non-GFP cells. Only relatively small differences were seen between the two cell populations and between the two experiments (Figure 2).

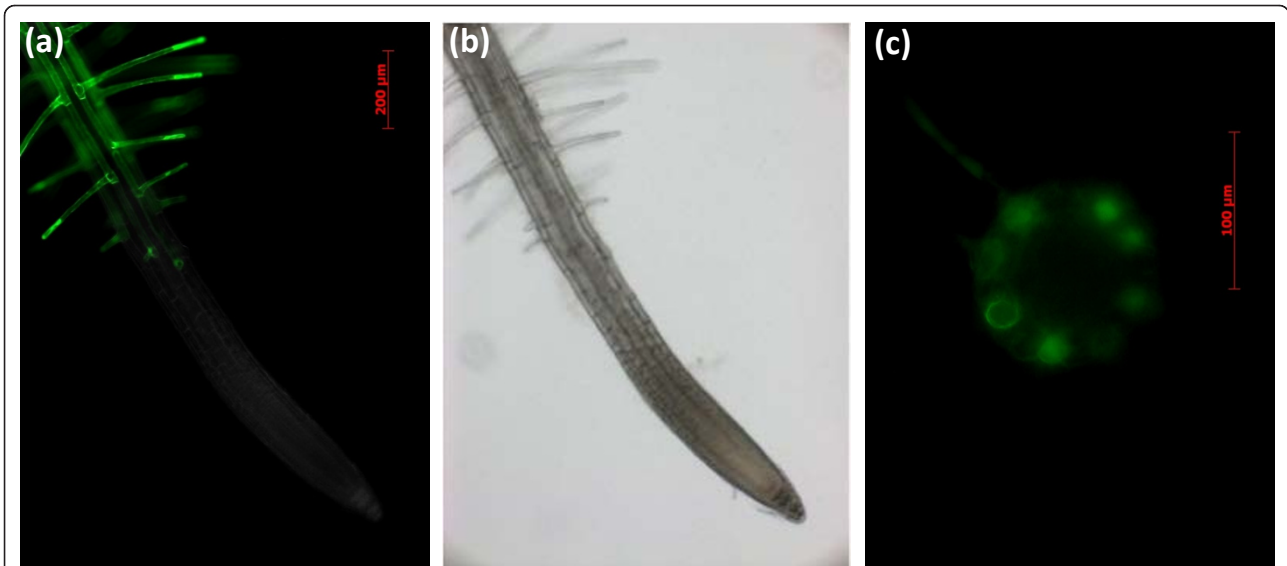


Figure 1 *pEXP7-GFP* expression pattern in roots of 5-day-old seedlings. (a, b) Root tip, (c) Cross-section in the root-hair zone. (a,c) GFP fluorescence; (b) bright field micrograph. GFP, green fluorescent protein.

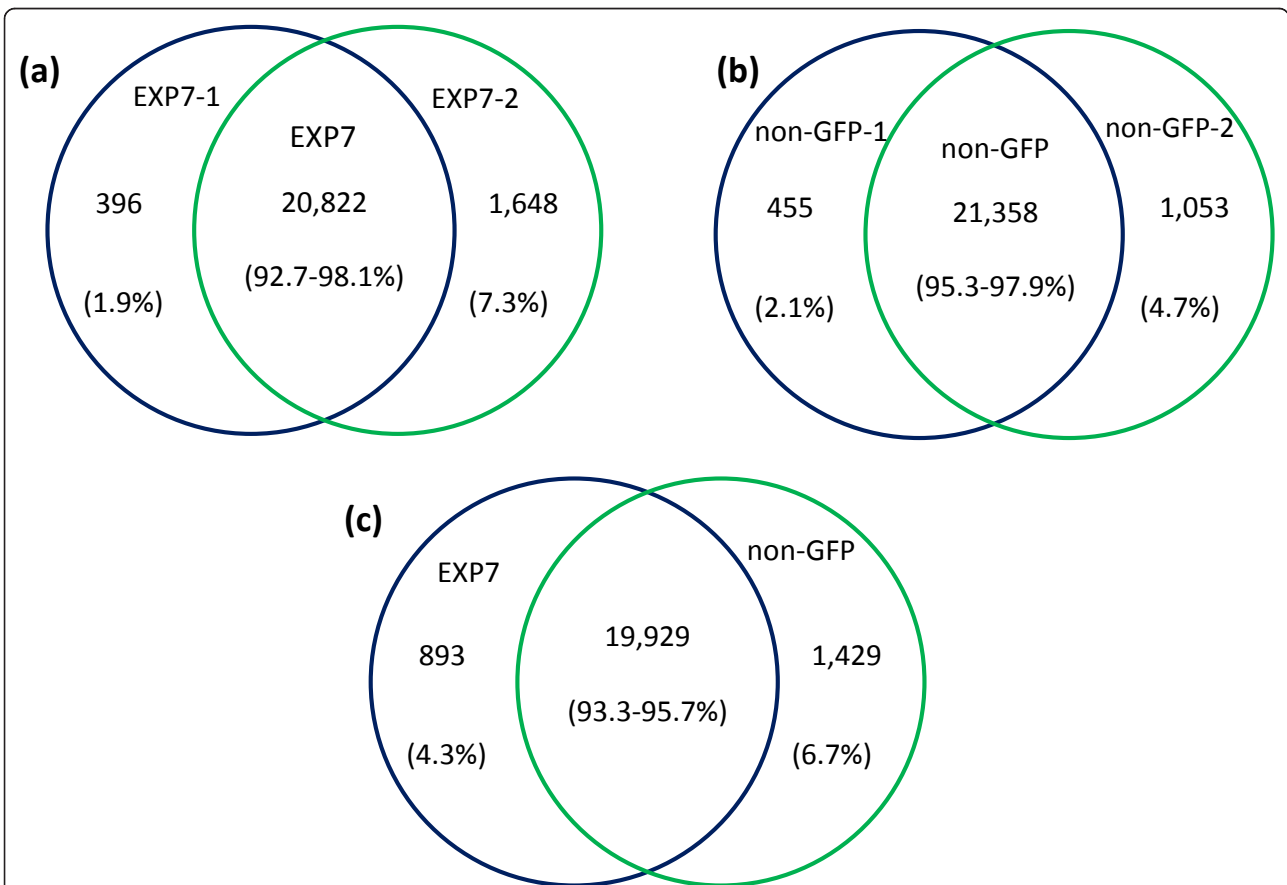


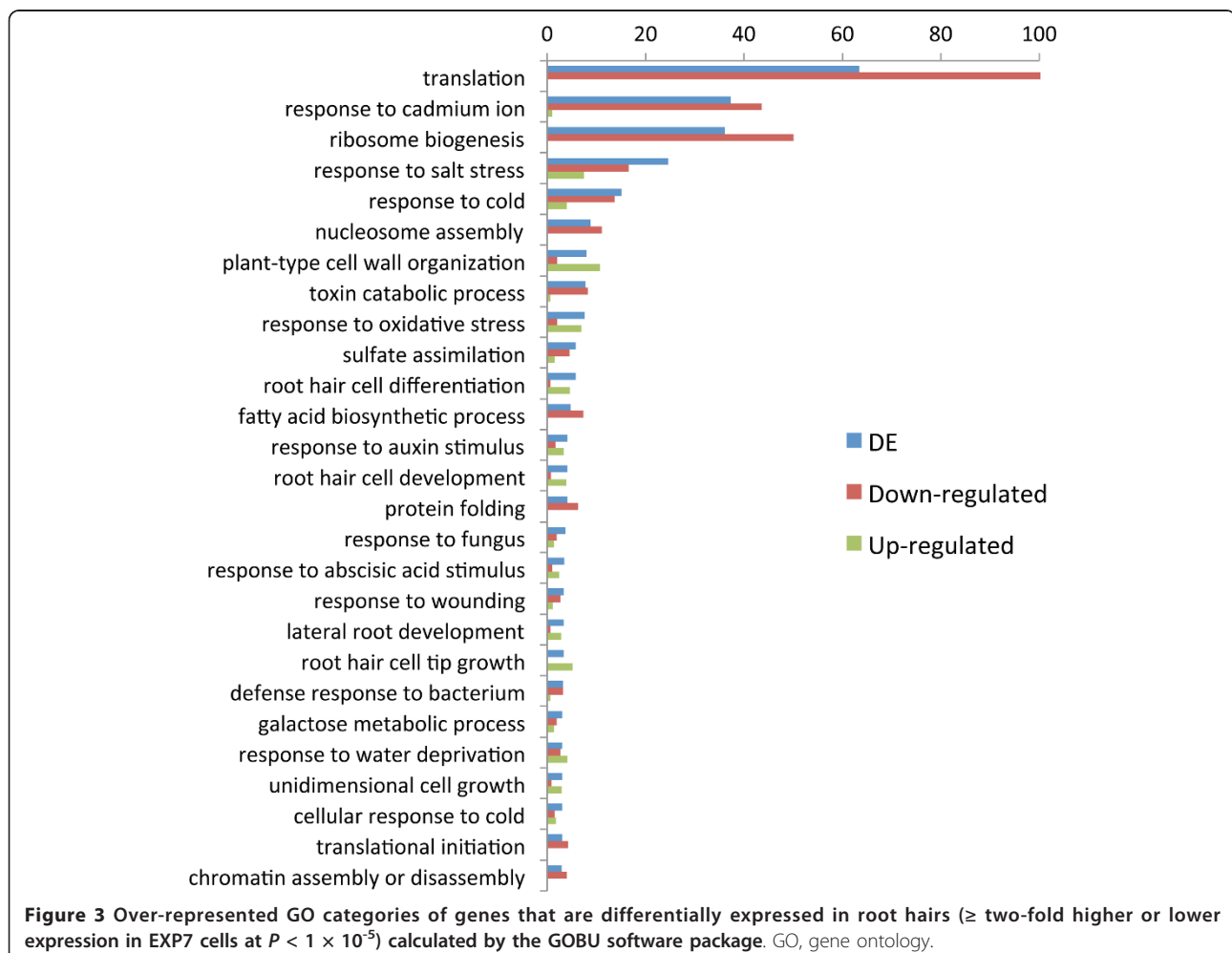
Figure 2 Identification of transcripts. (a,b) Venn diagrams of transcripts identified by at least five reads in two separate experiments in (a) EXP7 cells and (b) non-GFP cells. (c) Comparison of the identified transcripts in the two cell types. GFP, green fluorescent protein.

Differential gene expression between EXP7 and non-GFP Cells

While the numbers of genes defined as being expressed were comparable in the two protoplast populations, dramatic differences in abundance were seen for 1,617 transcripts ($P < 0.001$ in both experiments, $FC > 2$; see Additional file 3, dataset S2). Validation of RNA-seq results by quantitative reverse transcription (qRT)-PCR on 11 randomly chosen genes showed that there was a comparable dynamic range of and good agreement between the methods (see Additional file 1, Figure S2). A detailed list of functional categories identified as over-represented using the Gene Ontology Browsing Utility (GOBU) software package [15] is shown in Figure 3. As anticipated, high enrichment in EXP7 cells was seen for genes in the Gene Ontology (GO) categories 'plant-type cell wall organization', 'root hair cell differentiation,' and 'root hair cell tip growth'. In addition, genes from the categories 'response to salt stress' and 'response to oxidative stress' were highly enriched. Among the genes with markedly lower message levels in EXP7 cells, a large subset was

related to translation and ribosome biogenesis. In addition, genes involved in ribosomal RNA processing and genes encoding translation initiation factors had lower expression in EXP7 (see Additional file 3, dataset S2).

Epidermal cell fate in *Arabidopsis* is controlled by a complex interplay between cell-autonomous and non-cell-autonomous transcription factors. Consistent with current models of root epidermal-cell specification and differentiation, genes controlling cell fate and root-hair morphogenesis were for the most part differentially expressed between both cell populations. For example, expression of the *WER* gene was much lower in EXP7 protoplasts than in non-GFP cells, an observation that is consistent with the role of *WER* in suppressing the hair fate (see Additional file 3, dataset S2). After cell fate is determined, root-hair initiation starts with the formation of a dome-shaped structure at the basal end of the trichoblast. Transcripts of the basic helix-loop-helix (bHLH) -type transcription factor *RHD6* and of the Rho GTPase GDP dissociation inhibitor *SUPERCENTIPEDE* (*SCN1*), which both play crucial roles in root-hair initiation [16-18], were significantly



enriched in EXP7 cells. The mRNAs of several other genes with roles in root-hair elongation such as *LXR1*, *COW1*, and *RHD4* were also more abundant in EXP7 cells relative to non-GFP protoplasts. Genes with reportedly root hair-specific expression such as *EXP7* and *AGP3* showed about 800-fold higher abundance in EXP7 cells (see Additional file 3, dataset S2). It is interesting to note that some genes known to be crucial in root-hair elongation such as *RHD3* and *TIP1* showed similar expression levels in both cell types, indicative of important roles of these genes in other cell types.

The cell-expansion machinery in root hairs requires the activation of both carbon skeleton-providing and energy-providing processes. Transcripts from several genes involved in carbon metabolism, in particular in the tricarboxylic acid (TCA) cycle and gluconeogenesis, were highly enriched in root hairs. The high enrichment of transcripts encoding cytosolic phosphoenolpyruvate carboxykinase 1 (PCK1), which catalyzes the generation of phosphoenolpyruvate from oxaloacetate, and the NADP⁺-dependent malic enzyme 2 (ME2), which converts malate into pyruvate, indicate increased gluconeogenesis (Figure 4). This pathway might be activated to support the synthesis of glucose-6-phosphate as a precursor for cell-wall synthesis. Notably, we found that transcripts encoding proteins involved in oxidative phosphorylation (for example PCK1) also accumulated in EXP7 cells, probably necessary to fuel the reaction catalyzed. Transcripts derived from alcohol dehydrogenase *ADH1* were also highly enriched, probably to reoxidize the massive amount of NAD(P)H that is formed during the production of pyruvate and oxaloacetate (Figure 4).

Discriminative coexpression analysis links new genes to a root-hair elongation network

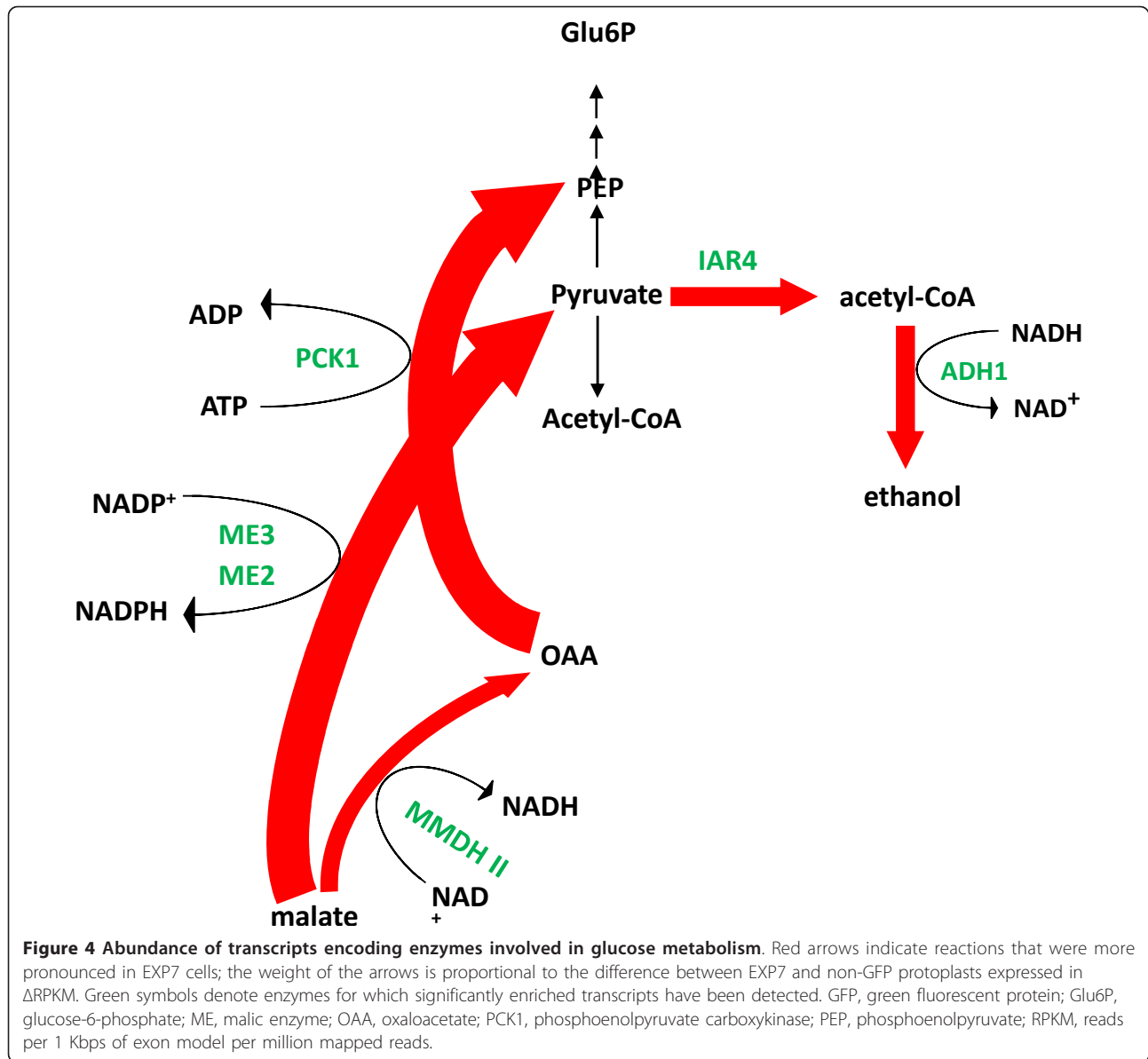
To identify candidate genes with potentially crucial roles in root-hair morphogenesis, we conducted a stringent coexpression clustering analysis of the 635 genes that had significantly higher expression level in EXP7 cells, using the in-house MACCU toolbox described previously [19]. Coexpression relationships between these genes were identified based on their Pearson's correlation coefficient being greater than or equal to 0.95 against a database of 111 array hybridizations (see Additional file 4, Table S1) taken from the NASCarrays database [20], which we selected for gene-expression signatures that discriminate processes associated with root-hair development. Clustering the 635 genes preferentially in EXP7 cells against a database of the selected arrays (see Additional file 4, Table S1), based on their pairwise coexpression with a cutoff of 0.95, generated a cluster of 98 genes (see Additional file 1, Figure S3). Many of these genes were shown in previous studies to be involved in root-hair formation: for example, *LXR1*, *COW1*, *RSL4*, *AHA*,

and *MRH9* [17,21-23, and several *RHS* (*ROOT HAIR SPECIFIC*) genes that carry the RHE (root hair element) consensus sequence in their promoters [24]. Notably, the cluster contained several genes with unknown function and genes that have not yet been associated with root-hair formation, providing a suite of putatively novel players in root-hair morphogenesis. It is also noteworthy that the GO categories 'response to oxidative stress' and 'oxidation-reduction process' were highly over-represented in the network, underlining the importance of redox processes for root-hair development.

To confirm the role of some genes in this cluster, we analyzed the phenotypes of 14 homozygous mutants harboring mutations in the coexpressed genes. We found that mutations in six of these genes caused alterations in the root-hair phenotype (Table 1; see Additional file 1, Figure S4). Homozygous mutants defective in the expression of At3g49960 formed significantly longer root hairs than did the wild type. Mutants harboring defects in the peroxidase At1g05240 and in the zinc transporter ZIP3 formed shorter root and irregular hairs when compared with the wild type. A similar root-hair phenotype was seen in homozygous mutants defective in the serine protease inhibitor (SERPIN) family protein CCP3 (CONSERVED IN CILIATED SPECIES AND IN THE LAND PLANTS 3). From the expression pattern and mutant phenotypes, a function for these genes in root-hair morphogenesis was deduced.

Whole root and cell type-specific alternative splicing

Using the in-house RACKJ software package, an algorithm specifically developed to analyze splice junctions and gene-expression levels from RNA-seq datasets, we investigated the four main types of alternative splicing in plants; intron retention, exon skipping (cassette exon), alternative 5' splice site (alternative donor), and alternative 3' splice site (alternative acceptor) at single cell-type resolution. A total of 83,615 and 101,942 splice junctions (SJs) were identified in EXP7 and non-GFP cells, respectively (Figure 5a; see Additional file 5, dataset S3). The percentage of SJ-containing genes was 64% in EXP7 and 71% in non-GFP cells (Figure 5e). Subsets of 8,842 genes and 11,058 genes produced alternative transcript variants in EXP7 and non-GFP cells, corresponding to 66.3% and 73.2% of all SJ-containing genes (Figure 5f). Considering all root protoplasts, 75.7% of the SJ-containing genes were alternatively spliced in roots, a number that is considerably higher than other estimates that have been published on alternatively spliced genes in *Arabidopsis* using RNA-seq [25,26]. We identified 4,940 alternative 5' and 2,859 alternative 3' splice-site events with at least two reads mapped to different starting positions and a minimum of five reads, respectively. In 2,439 cases, both donor and acceptor sites were different (Figure 5b). Generally, alternative splicing was

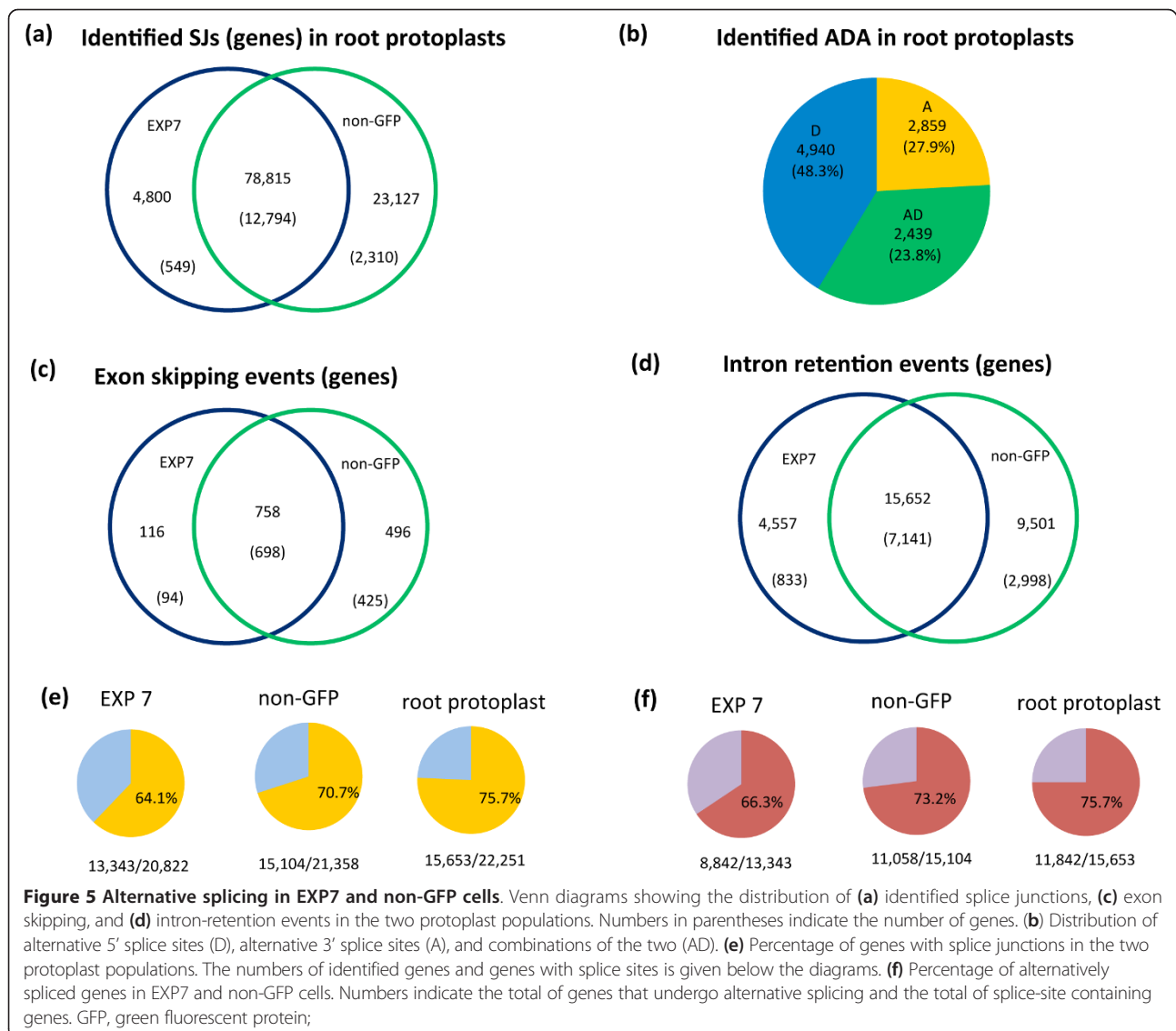


more complex in non-GFP cells than in EXP7 cells. For example, 1,254 and 874 exon-skipping events were seen in non-GFP and EXP7 cells, respectively. Similarly, 25,153 retained introns were identified in non-GFP cells and 20,209 in EXP7 cells (Figure 5c,d; see Additional file 6,

dataset S4). As anticipated from previous studies and in contrast to mammalian cells [27,28], we found that exon skipping was less common than other forms of alternative splicing, with the percentage of exon skipping estimated to be 2.4% of all alternative-splicing events.

Table 1 Phenotypes of mutants harboring defects in genes that are preferentially and coexpressed in root hairs

Locus	Gene	Molecular function	Mutant line	Phenotype
At3g49960		Oxidation-reduction process	Salk_119795C	Longer, more frequent root hairs
At1g05240		Peroxidase activity	Salk_10357C	Shorter, irregular, but more frequent root hairs
At2g32270	ZIP3	Zinc-transporter activity	Salk_008709	Shorter, irregular root hairs
At2g25240	CCP3	Serine-type endopeptidase inhibitor activity	Salk_076391; Salk_014163C	Shorter, irregular root hairs
At3g60330	AHA7	Hydrogen-exporting ATPase activity	SALK_042485	Reduced frequency of root hairs [43]
At5g60660	PIP2;4	Water-channel activity	Sail_1251_G12, Sail_535_D05	Longer root hairs [66]

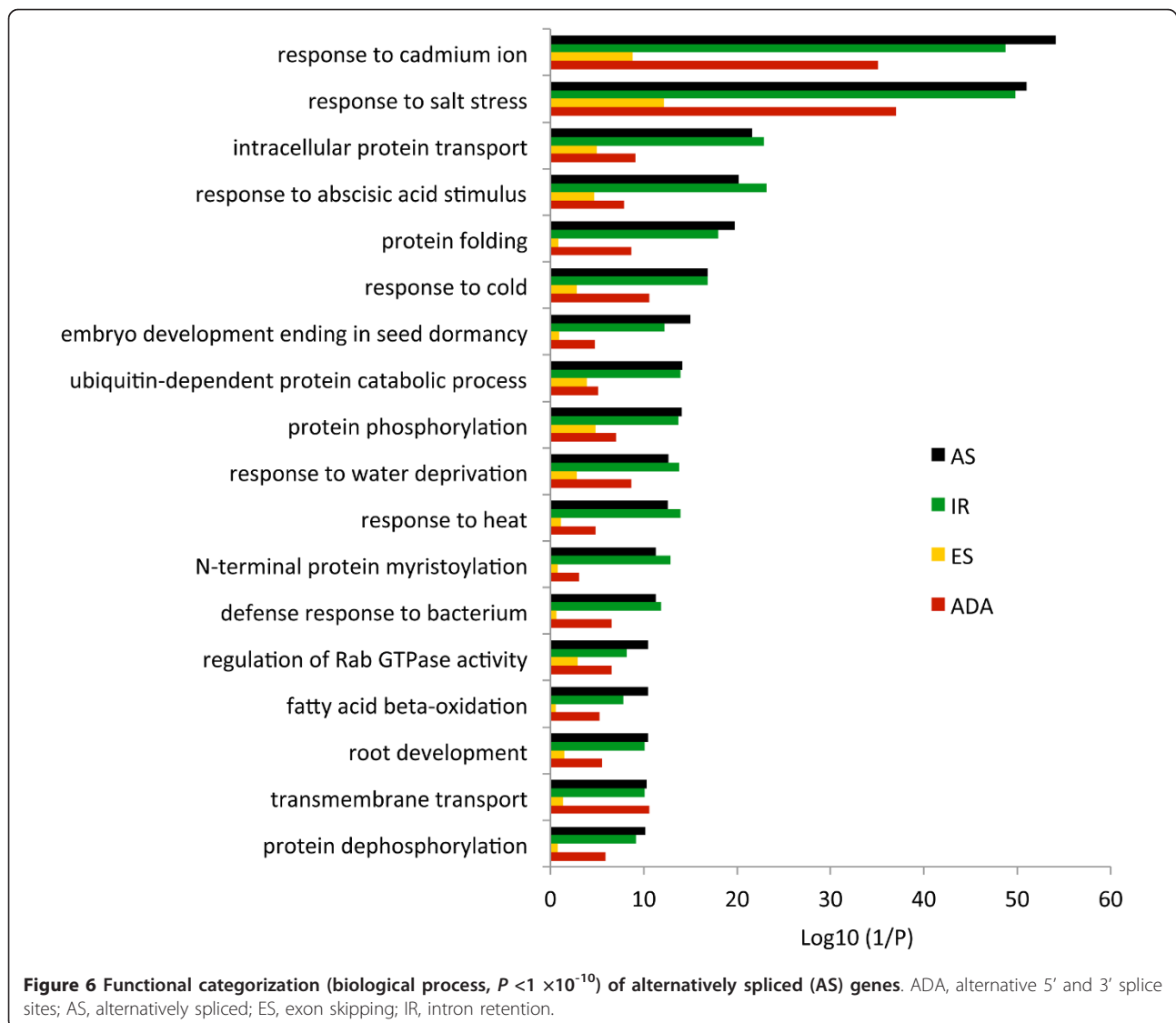


A comparison of the different forms of alternative splicing for alternatively spliced genes showed that for most over-represented GO categories ($P < 1 \times 10^{-10}$), intron retention was the major type of event, and exon skipping contributed the least to the total alternative-splicing events (Figure 6). Generally, intron retention was the most frequent form of alternative splicing. The GO category 'transmembrane transport', which was dominated by alternative 5'/3' splice-site selection, was the sole exception to this generalization.

Differential alternative splicing

An alternative-splicing event was considered to be cell type-specific when it was supported by five or more reads, a P -value of less than 0.05, and a difference of two-fold or greater in relative expression level between the two cell types. Using these criteria, we defined a

subset of 139 exon-skipping events as being significantly different between root hairs and non-GFP cells (differential exon skipping; DES) (Figure 7a). In this group, several genes were annotated as encoding proteins involved in pre-mRNA splicing, including the splicing factors RS41 and At2g16940, the ribosomal protein S8e family protein At5g06360, and the protein kinase AFC2. The products of several transcripts with differentially skipped exons were related to cell morphogenesis, such as the villin homolog VLN3, LONGIFOLIA 2, the spermine synthase ACL5, the leucine-rich repeat protein kinase family protein At3g21340, and MRH6. Several genes involved in vesicle transport also produced transcripts with differentially skipped exons (for example, *SYP51*, *RAB7A*, *VTI12*, *CASP*, *At1g50970*). Notably, for *PCK1*, transcripts with differential cassette exons were detected, which indicates a possible involvement of



alternative splicing in the regulation of carbon flux in root hairs.

We found 7,807 intron-retention features that were different between EXP7 and non-GFP cells (differential intron retention; DIR) (Figure 7b). A subset of 17 DIR and DES events was randomly selected for validation by qRT-PCR, all of which were found to have expression patterns that were consistent with the RNA-seq results (see Additional file 1, Figure S5). Intron retention was largely induced in EXP7 cells (that is, isoforms with retained introns were more prominent in EXP7 samples relative to non-GFP samples). Because intron retention frequently leads to the inclusion of premature stop codons, the higher number of transcripts with retained introns in root hairs is indicative of a potentially higher rate of unproductive splicing in this cell type. Among the genes with more than 10-fold suppressed intron-retention events (that is, where isoforms

with retained introns were less prominent in EXP7 samples relative to non-GFP samples, indicative of a high fidelity of the splicing process), several transcripts encode proteins that are associated with root-hair development or play important roles in cell-wall remodeling such as EXP7, IRE, MRH2, XTH5, XTH14, and the actin depolymerization factors ADF8, ADF9, and ADF11. In addition, several genes with root hair-specific expression were found to have reduced transcripts with retained introns in EXP7 cells (for example, *IRE*, *RIC1*, *REN1*, and the protein kinases At2g41970 and At5g16900). Notably, within this group of 257 genes with suppressed intron retention, 17 genes encoded protein kinases, suggesting an extensive cell type-specific regulation of protein phosphorylation by alternative splicing. It is interesting to note that whereas intron retention was markedly higher in EXP7 cells compared with non-GFP cells (65% up), this pattern was reversed when

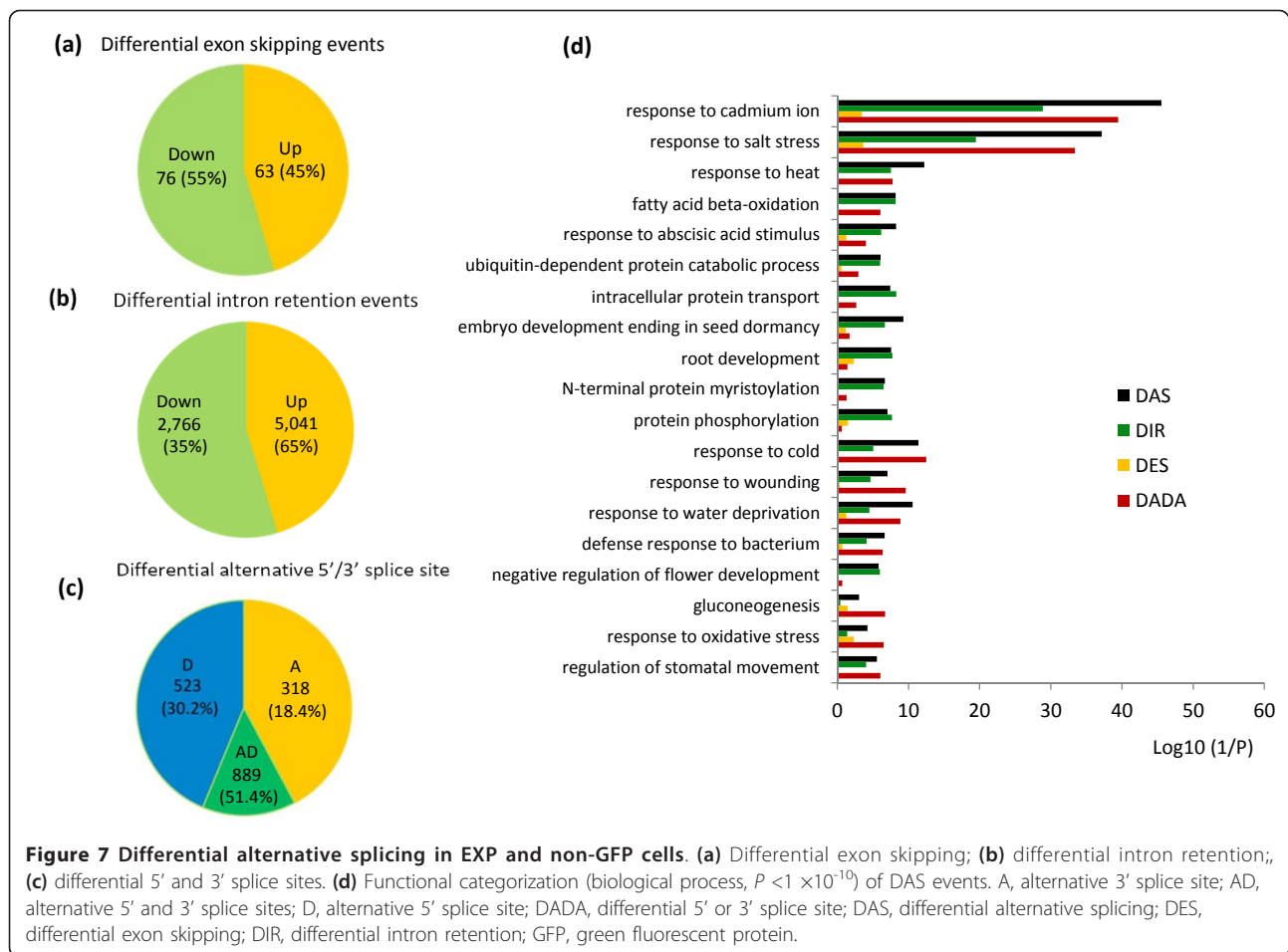


Figure 7 Differential alternative splicing in EXP and non-GFP cells. (a) Differential exon skipping; (b) differential intron retention; (c) differential 5' and 3' splice sites. (d) Functional categorization (biological process, $P < 1 \times 10^{-10}$) of DAS events. A, alternative 3' splice site; AD, alternative 5' and 3' splice sites; D, alternative 5' splice site; DADA, differential 5' or 3' splice site; DAS, differential alternative splicing; DES, differential exon skipping; DIR, differential intron retention; GFP, green fluorescent protein.

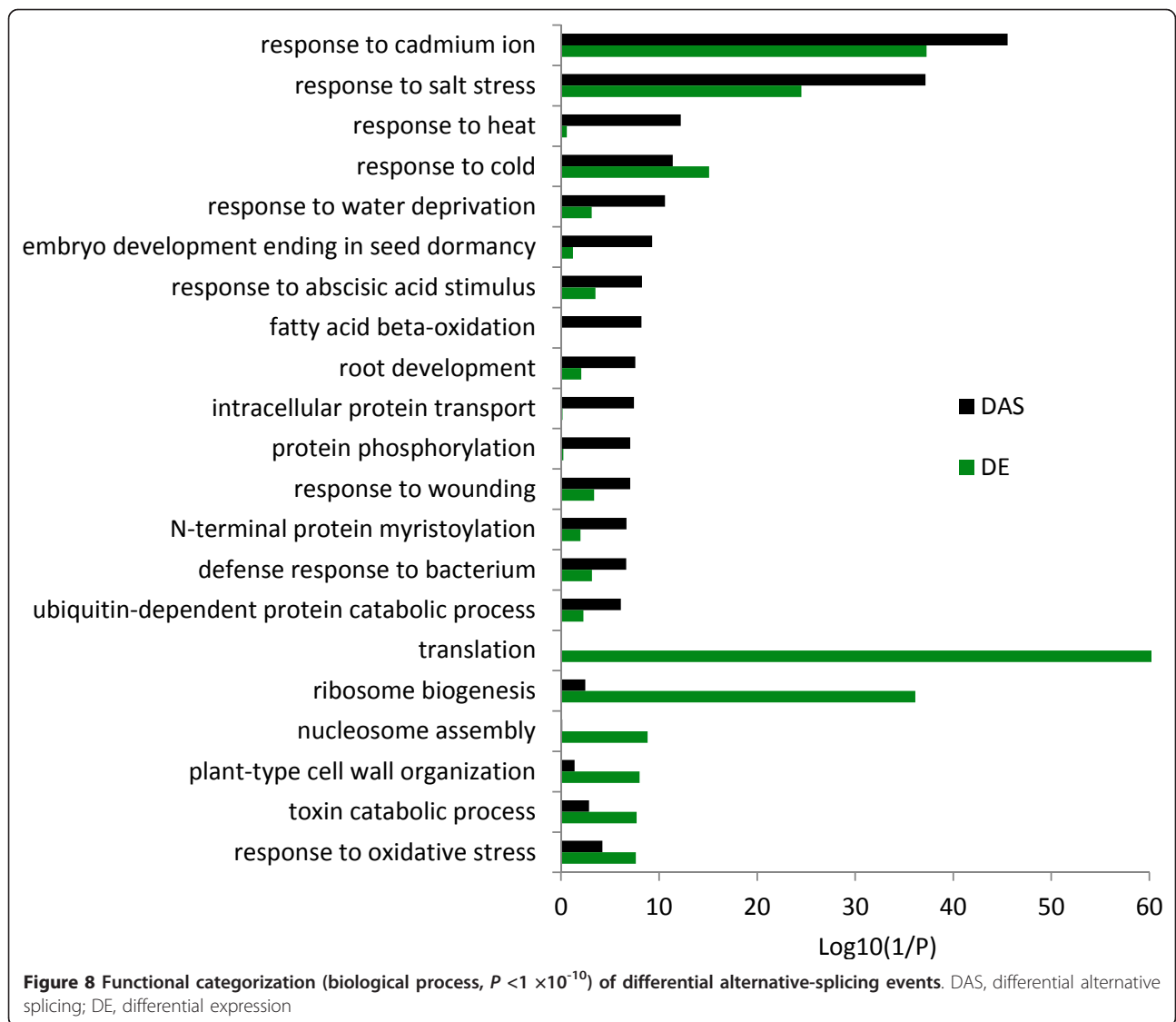
only genes from the coexpression network were considered (77% reduction in EXP7 cells). The number of genes that produced transcripts with differential alternative splice sites (differential alternative donor/acceptor; DADA) was 1,239 (1,730 events), the majority of which (51.4%) were due to both alternative 5' and 3' splice sites (Figure 7c). Interestingly, *REN1* and *IRE1* were among the genes that produced transcripts with DADA events, indicating complex regulation of these genes at the pre-mRNA level. In addition, several splicing factors or genes that encode splicing factor domain-containing proteins (for example, *At1g60200*, *RSP35*, and *SCL33*) were included in this group. Notably, *RSP35* transcripts were less abundant in EXP7 cells, indicating regulation at both the transcriptional and post-transcriptional level, and a possible function in the establishment of cell type-specific splicing patterns.

As might be anticipated from the distribution of the four forms of differential alternative splicing (DAS), DIR and DADA events were predominant in all over-represented functional categories ($P < 1 \times 10^{-6}$) (Figure 7d). Notable exceptions were the categories 'response to oxidative stress' and 'gluconeogenesis,' in which DADA was most

prominent. A comparison of differentially expressed genes with those that underwent differential alternative splicing showed that in several over-represented GO categories, the differential expression of genes was not associated with DAS events (for example, 'ribosome biogenesis,' 'translation,' nucleosome assembly,' and 'cell wall organization'), whereas other categories were dominated by DAS (for example, 'protein phosphorylation' and 'intracellular protein transport') (Figure 8). The latter finding matches the observation that for several genes involved in protein phosphorylation, intron retention was largely repressed in EXP7 cells. Differentially expressed genes and genes that formed transcripts that were alternatively spliced neither showed a large overlap, nor were the two processes mutually exclusive. For a comprehensive overview of the differences in splicing pattern between the two cell types, see Additional file 7 (dataset S5).

Identification of root-hair proteins

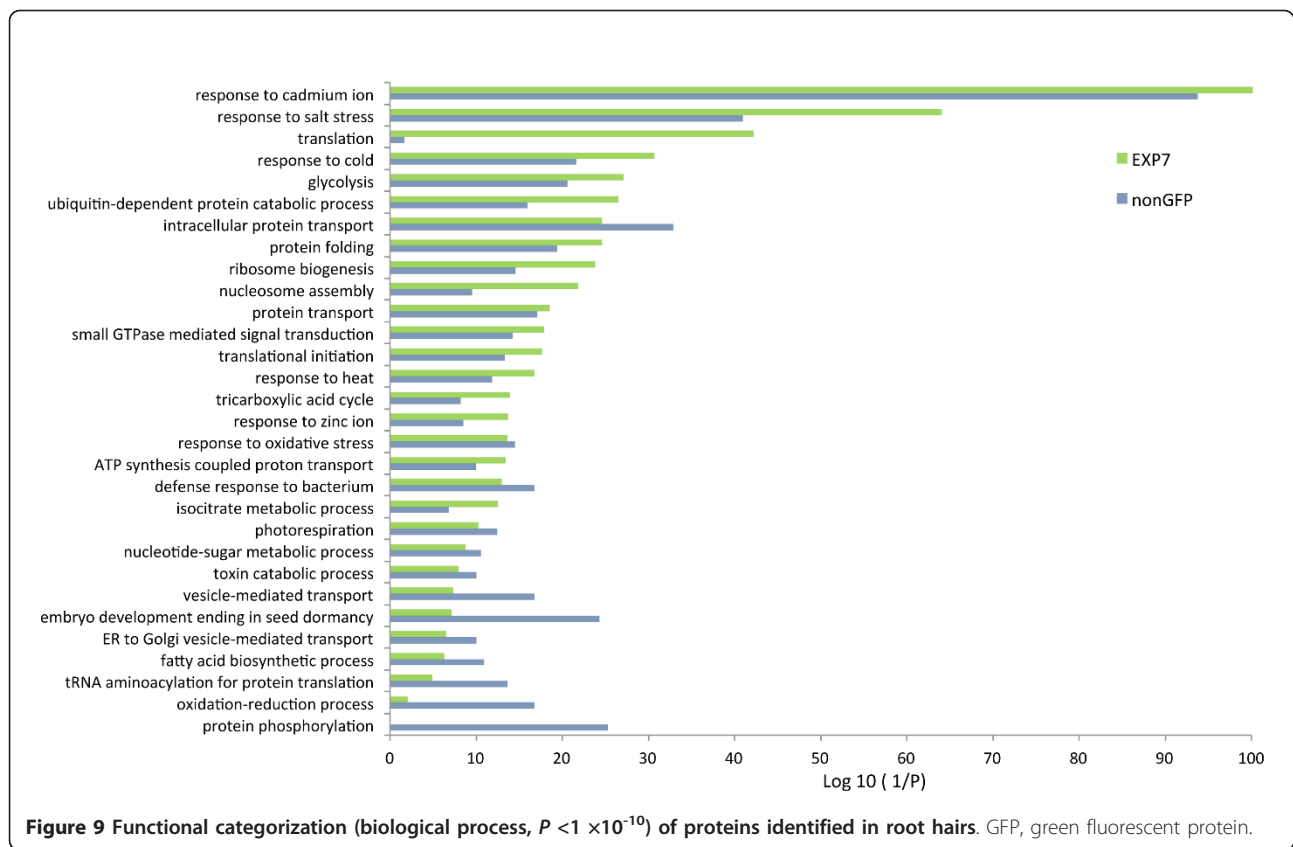
A high-density root-hair proteome map was generated by analyzing the proteomic profiles of the two protoplast populations by means of liquid chromatography/tandem



mass spectrometry (LC/MS-MS) on a mass spectrophotometer (LTQ Orbitrap Velos, Thermo Fisher Scientific Inc., Rockford, IL, USA). In total, 792,765 spectra from 50,415 peptides were detected in root protoplasts, corresponding to 12,492 proteins (see Additional file 1, Figure S6; see also data available in dataset S6 [29]). A comparison between all identified proteins and the TAIR10 proteome annotation found comparable molecular weight and isoelectric point distributions, indicating almost complete coverage of expressed proteins (see Additional file 1, Figure S7). An overview of the over-represented GO categories for the proteins in EXP7 and non-GFP cells is given in Figure 9. A large subset of the peptides identified was found to have post-translational modifications (see Additional file 4, Table S3).

For further analyses, only those proteins that were identified in two independent experiments or by at least

two peptides in one of the two experiments were considered. To compare the protein abundance between the two cell populations, we used the normalized spectral abundance factor (NSAF) method with a confidence limit of 95% [30,31]. A subset of 33 proteins identified as robustly and exclusively expressed in EXP7 cells was referred to as root hair-specific proteins (see Additional file 4, Table S4). We defined 96 proteins as differentially expressed, with a cutoff of 4.40-fold and 0.167-fold for proteins with increased and decreased abundance in root hairs, validated by power analysis with $\beta = 0.80$ (see Additional file 4, Table S5). Consistent with the function of root hairs, proteins involved in energy metabolism and transport showed increased expression in EXP7 protoplasts. Interestingly, several histone-family proteins and nucleosome core components also had higher abundance in EXP7 protoplasts, possibly indicating that chromatin



modifications are crucial in controlling transcriptional activity in root hairs. In line with the assumption that translation is decreased in root hairs, several proteins related to tRNA aminoacylation and aminoacyl-tRNA ligase activity showed lower expression in EXP7 cells relative to non-GFP protoplasts.

Defining the root-hair interactome

A root-hair interactome that comprises all protein-protein interactions (PPIs) of the EXP7 proteome was constructed based on confirmed interactions provided by the *Arabidopsis* Interactome Mapping Consortium [32] and by The *Arabidopsis* Information Resource (<http://www.arabidopsis.org/portals/proteome/proteinInteract.jsp>). In total, 395 non-reciprocal interactions between proteins plus 77 confirmed self-interactions could be deduced from these data (see Additional file 4, Table S6). Experimentally verified PPIs may not have occurred *in vivo* if the proteins were located in separate tissues. Our analysis verified colocalization for a comprehensive subset of the confirmed PPIs, allowing more accurate assessment of functional associations. A PPI network based on this database was comprised of one large and several smaller sub-clusters, plus a set of protein-protein pairs that interact only with one partner (see Additional file 1, Figure S8). Smaller sub-clusters were comprised of proteins related to protein binding/folding,

intracellular protein transport, calcium signaling, and mitochondrial electron transport, processes that are associated with tip growth. Several root hair-specific proteins or proteins that were differentially expressed in EXP7 cells have interacting partners within this network. For example, the actin-binding formin homology protein FH5 interacts with ACTIN-12 (ACT12), which is connected with several other actins and the actin depolarizing factor ADF3. This sub-cluster is further connected with several genes related to cell-wall organization and intracellular protein transport, with the clathrin At3g08530 as a central node. Remarkably, the abundance of mRNA was discordant with that of proteins. For example *ACT2/DER1*, important for cell elongation and root-hair formation, showed higher transcript abundance in EXP7 cells compared with non-GFP cells, but no differential expression at the protein level. Similarly, *FH5* transcript levels did not differ between the two cell types, despite the abundance of FH5 protein being 5.7-fold higher in EXP7 cells. Notably, no corresponding transcript was detected for ACT12.

Concordance of protein and transcript abundance

In contrast to what has been expected from the relatively low correlation of root-specific proteins with their cognate transcripts (see Additional file 4, Table S4), the general concordance of protein and mRNA abundance was

remarkably high. Considering all transcript-protein pairs for which the transcript was defined as significantly enriched in EXP7 cells and the cognate protein was identified and quantified (129 transcript-protein pairs), a correlation of $r^2 = 0.65$ was found, which is comparable with other quantitative mRNA/protein comparisons in *Arabidopsis* roots and *Chlamydomonas* cells (Figure 10a) [33,34]. The correlation was similar when calculated for all quantified proteins with higher abundance in EXP7 cells relative to non-GFP cells for which the corresponding transcript was differentially expressed at $P < 0.05$ (323 protein-transcript pairs; Figure 10c). The abundance of mRNA and protein was highly discordant when transcript-protein pairs with decreased abundance in EXP7 cells were considered (435 transcript-protein pairs; Figure 10b). Similarly, no correlation was seen when

protein-transcript pairs were selected in which the protein was less abundant in EXP7 cells (453 protein-transcript pairs; Figure 10d). Together the results suggest that while generally upregulation of gene activity was associated with higher abundance of both mRNA and protein, decreased abundance of a gene product was not correlated with a similar trend in the cognate partner.

Discussion

Integration, analysis, and interpretation of data from different 'omics' levels remain a major challenge in systems biology. Our analysis provides a comprehensive, multilayered reference map of the *Arabidopsis* root hair without bias caused by amplification, tissue heterogeneity, or factors introduced when independently collected data are merged. This dataset catalogues the components of root-hair cells

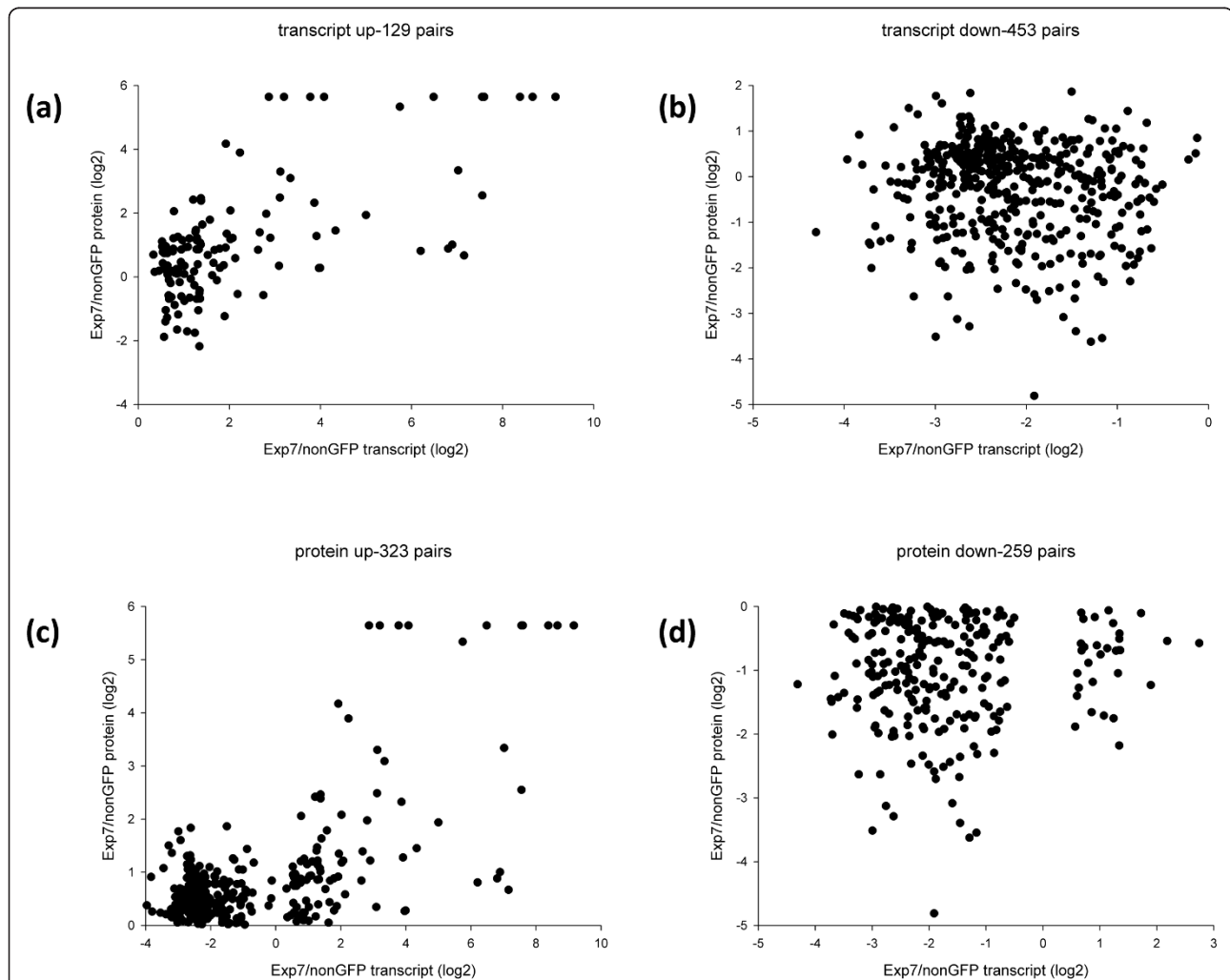


Figure 10 Concordance between differences in the abundance of mRNA and its encoded protein. (a-d) Correlation between protein and transcript fold-differences between EXP7 and non-GFP cells for transcripts with (a) higher or (b) lower abundance in EXP7 cells, and for proteins with (c) higher or (d) lower abundance in EXP7 cells. Concordance was calculated from 582 transcript-protein pairs derived from 1,774 quantified plus 33 root hair-specific proteins and 1,850 differentially expressed ($P < 0.001$) transcripts. GFP, green fluorescent protein.

and may aid in deciphering the molecular machinery that orchestrates cell-differentiation processes and thus deepen our understanding of the regulation of these processes. Our dataset complements and extends previous transcriptomic and proteomic studies in *Arabidopsis* using either mutants with defects in root-hair morphogenesis and subsequent transcriptomic or proteomic profiling, or FACS-based approaches, or combinations of the two [9,11-13,35,36]. We compared the root-hair (EXP7) datasets with data obtained from analyzing all root tissues except root hairs (non-GFP), allowing disclosure of differentially expressed genes and proteins without distorting data by using a separate GFP reporter line for non-root-hair cells, which might differ in terms of signal strength or in the spatial expression of the reporter chimera. From a total of 22,251 identified genes, 7.3% were classified as differentially expressed between root hairs and non-GFP cells. From the 12,492 proteins identified in roots, 2,447 were detected in EXP7 cells. From these proteins, a subset of 129 (5.3%) proteins accumulated differentially. Taking into account the different methods for determination and for calculating thresholds for differential expression, these numbers are comparable, providing a robust database for comparing differential gene expression at the transcript and protein level.

A subset of 93 genes was expressed specifically in root hairs (reads per 1 Kbps of exon model per million mapped reads (RPKM) <1 in non-GFP cells). Contained in this set were several RHE motif-containing genes and other genes that have been associated with root-hair differentiation such as *MRH6*, *LRX1*, and *DER4*. As anticipated, most root hair-specific genes encode cell wall-related proteins such as extensins, arabinogalactan proteins, xyloglucan endotransglucosylases, and pectinesterases. In addition, several genes with putative functions in signaling pathways, such as transcription factors, protein kinases, and ROP GTPase-related genes, were specifically expressed in root hairs. Small GTP-binding proteins play crucial roles in signal transduction and cytoskeletal organization. Three genes involved in the regulation of Rho-like small GTPases (ROPs) were within this group. RHO GUANYL NUCLEOTIDE EXCHANGE FACTOR 4 (ROPGEF4; also known as RHS11) was shown to regulate ROP11-mediated cell-wall pattern in xylem cells. In roots, ROPGEF4 is crucial for the control of root-hair elongation under Pi-deficient conditions [19]. ROP-INTERACTIVE CRIB MOTIF-CONTAINING PROTEIN 1 (RIC1) is a key player in the signaling network dictating the shape of leaf pavement cells by controlling the organization of cortical microtubules [37]. Notably, RIC1 is coexpressed with another cell-specific gene, the protein kinase At5g18910. The third GTPase-related gene in this group is the RAB GTPASE HOMOLOG H1D (RABH1D). RABH1D is coexpressed with the AGC kinase INCOMPLETE ROOT HAIR ELONGATION (IRE), which has specific expression in

root hairs and is essential for root-hair elongation [38]. Given the crucial roles of these proteins in cell shape, the root hair-specific expression pattern highlights the important function of the Rho-signaling network in root-hair differentiation. and implicates ROPGEF4, RIC1, and RABH1D as key components.

Large differences between EXP7 and non-GFP cells were seen in the abundance of transcripts coding for mineral-nutrient transporters. Root hairs provide a large proportion of the interface of the plant and the rhizosphere, thus, a higher abundance of transcripts encoding proteins involved in the uptake of nutrients from the soil solution in root hairs would be expected. Transporters with clearly defined roles in nutrient acquisition from the soil solution such as *SULTR1;1* and *NRT2;1* were highly expressed in EXP7 cells and virtually absent in non-GFP cells, suggesting specific roles of the encoded proteins in root hairs (Table 2). However, we cannot exclude that these genes are highly enriched in a small subset of tissues that is masked by the current approach. Interestingly, the proton ATPase *AHA7* was specifically expressed in root hairs. We previously showed that *AHA7* is upregulated by iron deficiency, but is not important for rhizosphere acidification, a process that aids in the acquisition of iron [23]. Rather, we found that *AHA7* is crucial for the formation of root hairs, a process that is affected by iron deficiency [39]. Regulation of the extracellular pH is crucial for cell-wall expansion in root hairs and pollen tubes [40,41]. Thus, the strong enrichment of *AHA7* transcripts in EXP7 cells might indicate a root hair-specific function of *AHA7*, possibly oscillatory acidification of the apoplast of the root-hair tip [40].

Transcriptionally coordinated genes often participate in a common function. We used a selective coexpression approach to identify previously unrecognized genes important in root-hair differentiation, using genes that were preferentially expressed in EXP7 as an input for network construction. Coexpression networks are usually assembled by clustering differentially expressed genes against a database comprising microarray experiments

Table 2 Abundance of transcripts encoding mineral-nutrient transporters in EXP7 and non-GFP cells

Locus	Transporter	RPKM	
		EXP7	non-GFP
At1g12110	NRT1;1 (NITRATE TRANSPORTER 1.1)	59.7	19.8
At1g08090	NRT2;1 (NITRATE TRANSPORTER 2.1)	37.8	0.3
At3g45060	NRT6;2 (HIGH AFFINITY NITRATE TRANSPORTER 2.6)	536.6	75.2
At5g43350	PHT1;1 (PHOSPHATE TRANSPORTER 1;1)	89.0	11.8
At4g08620	SULTR1;1 (SULPHATE TRANSPORTER 1;1)	19.0	0.0
At3g60330	AHA7 (H(+)-ATPASE 7)	224.7	1.4

Abbreviations: GFP, green fluorescent protein.

that cover a wide range of tissues and conditions, often leading to the inclusion of unrelated genes [42]. To strengthen the 'guilt by association' approach, we selected microarray experiments based on a suite of genes with confirmed function in root-hair development. For five genes in this network, mutant analysis identified novel roles in tip growth. Expression of At3g49960 was found to be upregulated by nutrient deprivation, and this upregulation was dependent on the NADPH oxidase RHD2 [43]. Further, At3g49960 was found to be preferentially expressed in root-hair cells, and this expression pattern was associated with higher H3K4me3 and lower H3K27me3 levels in root-hair cells relative to non-root-hair cells [44]. Gene activity of the peroxidase superfamily protein At1g05240 was detected preferentially in root hairs both at the transcript and protein level. Root hair-specific expression of the protein and a high transcript level that was approximately 650 times more abundant in EXP7 cells than in non-GFP cells suggest an important function of At1g05240 in root-hair morphogenesis, which is yet to be characterized. Mutants defective in this gene formed shorter and irregular hairs compared with the wild type (see Additional file 1, Figure S4). Another protein not previously associated with root-hair development was the Zn-responsive metal transporter ZIP3 [45]. Homozygous *zip3* mutants formed shorter root hairs, indicative of a role of Zn in tip growth. Another Zn-regulated gene, *IRT2*, was present in the network generated in the present study and also in the gene regulatory network reported by Bruet *et al.* [12], supporting a role for Zn or other transition metals that might be transported by ZIP3 and IRT2 in root-hair elongation. CCP3 is a serine protease inhibitor that has a predicted ancestral ciliary function and has been maintained in non-ciliated plants, probably to perform novel functions [46]. In addition, we confirmed roles for AHA7 and PIP 2;4 in root-hair development.

Alternative splicing generates diverse transcript isoforms from a limited number of protein-coding genes, leading to proteins that potentially possess unique functions. Alternatively, in combination with nonsense-mediated decay (NMD), unproductive mRNA splicing can lead to the synthesis and subsequent degradation of non-functional transcripts, thereby providing an alternative route of gene regulation (regulated unproductive splicing and translation; RUST) [25,47,48]. In plants, the most abundant types of alternative splicing are intron retention and alternative donor/acceptor [25,26,49], and the question as to whether alternative splicing, on a global scale, is of physiological significance or is due to 'noisy' splicing of pre-mRNA is still open to debate. FACS-based techniques allowed for detailed investigations into cell type-specific gene expression; however, information regarding differences in splicing patterns between different organs, tissues, or cell types in plants is rare. Although our data do not directly

address the biological relevance of alternative splicing, several lines of evidence makes it reasonable to assume that alternative splicing represents an important, but, in plants, largely unexplored layer of gene regulation. First, alternative splicing is less complex in root hairs compared with non-GFP cells. Assuming a function of alternative splicing that exceeds stochastic fluctuations, it can be assumed that the population of pre-mRNAs subjected to alternative splicing varies with the structure and function of the cell type, and is more intricate when several cell types are analyzed together. Second, intron retention was largely induced in EXP7 cells, but was strongly repressed for a subset of genes, many of which encode proteins crucial to root-hair morphogenesis. Unproductive alternative splicing has been shown to control the number of physiological transcripts in plants [50]. Splicing accuracy can provide a mechanism to fine-tune gene expression, adjusting protein activity to the desired level. Importantly, and in line with this explanation, we found that intron retention was largely repressed in genes comprising the coexpression network, which probably contain nodes with relevant functions in root-hair development. Thus, transcripts with crucial functions in a certain cell type or functional context seem to be more likely to be spliced with high fidelity. This was evident for a number of genes encoding protein kinases and phosphatases, suggesting that protein phosphorylation is largely regulated at the post-transcriptional level.

Although a comprehensive picture of the transcriptional profile of *Arabidopsis* root hairs is emerging from various approaches, the root-hair proteome has been much less explored. A recent study documented 1,363 proteins expressed in root hairs [13]. The current survey extends this list to 2,447 root-hair proteins, 96 of which accumulated differentially between EXP7 and non-GFP cells, and 33 of which were specifically expressed in root hairs. The overlap of our study and the previous one is represented by a subset of 935 proteins, indicating a high correlation between the two surveys. Of the 33 root hair-specific proteins, At1g05240 has been identified in the coexpression cluster as being crucial for root-hair elongation (see Additional file 1, Figure S4). The putative cysteine proteinase At3g43960 was one of eight loci previously identified as being involved in root-hair elongation under phosphate-deficient conditions from a coexpression network of phosphate-responsive genes [19]. In the same screen, another root hair-specific protein, the EF-hand-containing At1g24620, was found to be crucial for root-hair elongation. Although these findings support the assumption that the root hair-specific proteins identified here are crucial in root-hair morphogenesis, it is interesting to note that many proteins in this group were not associated with differential expression of their cognate transcript, supporting the assumption that

the transcript level is an insufficient proxy for protein abundance in a certain cell type.

With few exceptions, the proteins with preferred expression in EXP7 cells have not been previously associated with root-hair formation, but for many of them, a putative function in specific processes related to tip growth can be inferred from related proteins with defined functions. For example, the cell membrane-anchored formin homology protein FH5 was among the proteins that were strongly enriched in root hairs. Formins are actin-nucleating proteins with key roles in root-hair formation and pollen-tube growth. FH5 drives tip growth in pollen tubes by stimulating apical actin assembly [51]. Ectopic expression of another FH family protein, FH8, resulted in severely deformed root hairs and formation of multiple bulges [52] indicating a crucial function of FH8 in root-hair initiation and elongation; a similar function may be deduced for FH5. Another member of this family, FH2, for which no specific function has been described, was exclusively found in root-hair cells (see Additional file 4, Table S4). Transcript levels for both FH2 and FH5 were similar in EXP7 and non-GFP cells, indicating post-transcriptional regulation of protein abundance.

In addition to alternative splicing, post-translational modification (PTM) of proteins contributes to the functional diversity of the proteome. For many of the identified proteins, extensive modifications were detected in our study. Similar to a previous report [53], most of the modified peptides were either carbamidomethylated at cysteine, or oxidized at methionine. Oxidation of methionine, if occurring *in vivo*, may have functional implications as a redox switch. Acetylation was the second most pronounced peptide modification, including acetylation of the N-terminal α -amine of peptides and acetylation of the protein N-terminus, as well as acetylation of lysine, neutralizing the positive charge of this amino acid. Protein acetylation plays a crucial role in transcription regulation [54]. Recent studies showed that lysine acetylation acts as a regulatory switch in both human and prokaryote cellular metabolism by altering protein activity or stability [55,56]. Of particular interest is the regulation of the reversible phosphorylation of proteins. Our analysis shows that many protein kinases and phosphatases are preferentially expressed in root hairs and are robustly regulated by cell type-specific intron retention, indicating extensive regulation of protein phosphorylation at the post-transcriptional level. The pronounced effect of DAS on protein phosphorylation was in contrast to several over-represented GO processes that were associated with differential gene expression, but only to a minor extent with DAS (Figure 8). It is possible that, in general, processes that require fast regulation and the integration of several internal and external signals are more affected by DAS, whereas 'molecular machines' related to the

assembly of biological structures are chiefly regulated by differential expression.

Conclusions

Taken together, the transcriptome/proteome comparison at single cell-type resolution sets the stage for an integrative, functional interpretation of large-scale datasets. The central dogma of molecular biology proposes a highly correlated abundance of mRNA and protein. As anticipated from other integrative studies, we found that transcript and protein abundance were not always concordant, making analysis of these different levels of regulations mandatory for a comprehensive understanding of the processes that drive cell differentiation. It is obvious from the present investigation that cell differentiation is an integrated readout of several levels of regulation, including transcription, mRNA processing, translation, and PTMs of protein, levels that are not necessarily congruent with each other. This and other studies suggest that an understanding of processes related to cell differentiation and function requires study of all of these layers to deduce key regulatory circuits, and a holistic understanding of concerted interactions of several system components that act together as molecular machines in metabolic and regulatory pathways. The current compendium of proteins, transcripts, and splicing variants provides a plethora of candidates for further studies into the molecular mechanisms that drive root-hair differentiation.

Materials and Methods

Plant material and growth conditions

The accession Col-0 (*Arabidopsis* Biological Resource Center (ABRC), Ohio State University, Columbus, OH, USA) was used in this study. Transgenic plants carrying the *pEXP7-GFP* reporter construct were used (obtained from Daniel Cosgrove, Penn State University, Pennsylvania, PA, USA). Seeds were sterilized and plated on top of a nylon mesh (pore size 100 μ m) that was placed on Petri dishes. The plates were kept for 1 day at 4°C in the dark, and then grown as described previously [19]. Plants 5 days old were used for the experiments.

Bulk root protoplasting and cell sorting

Roots were cut off about 10 mm from their tip, and immersed into a solution (referred as solution B; a mixture of solution A (10 mmol/l KCl, 2mmol/l MgCl₂, 2 mmol/l CaCl₂, 0.1% BSA, 2 mmol/l MES, 600 mmol/l mannitol, pH 5.5) +1.5% cellulose (Yakult Honsha, Tokyo, Japan), and 0.1% pectinase (SigmaAldrich, St Louis, MO, USA)]. The dissected roots were incubated for 2 hours in solution B in the dark at room temperature with agitation. Protoplasts were collected and filtered through a 70 μ m cell strainer. The cell suspension was then separated by centrifugation (52 g) for 10 minutes at

4°C. The supernatant was aspirated, and the cell pellet was re-suspended in solution A. The cell suspension was then filtered through a 40 µm cell strainer and kept on ice.

GFP-expressing cells were isolated in a fluorescence-activated cell sorter (Moflo XDP; Beckman Coulter Inc., Brea, CA, USA), which was equipped with a cooling device and fitted with a 100 µm nozzle, at a rate of 7,000 to 8,000 events per second at a fluid pressure of 25 psi. Protoplasts from wild-type plants were used as a negative control for establishing sorting criteria. GFP-positive (EXP7) cells were selected by their emission intensity in the green channel (approximately 530 nm) above negative controls. Protoplasts from non-GFP-expressing cells were collected in the same runs. Cells were sorted directly into liquid nitrogen and stored at -80°C for subsequent RNA and protein extraction.

Library preparation and RNA sequencing

Total RNA was extracted (RNeasy Mini Kit; Qiagen Inc., Valencia, CA, USA). Paired-end cDNA libraries were constructed from 5 µg of total RNA from EXP7 and non-GFP cells, with insert sizes ranging from 400 to 500 bp using a commercial kit (TruSeq RNA Sample Preparation Kit; Illumina Inc., San Diego, CA, USA). The paired-end libraries were sequenced for 100 to 150 bases **on a single lane per sample of a genome analyzer** (GAIIx platform; Illumina). The RNA-seq data collection followed published protocols [57].

Computing differentially expressed genes

Paired-end reads generated from EXP7 and non-GFP samples were mapped to the TAIR10 genome using the BLAT program [58]. The RACKJ (Read Analysis & Comparison Kit in Java; [59]) software package was used to filter mapped reads so that a pair of reads was retained if both reads showed an overlap within the same gene region. A transcript was defined as present when it was detected with at least five reads in each of the two experiments within samples from either cell type (EXP7 or non-GFP). RPKM values were then computed as described previously. [57]. For each gene in each repeat, a two-sided *P*-value was calculated by the following Z-statistic under the standard normal distribution:

$$Z = \frac{p_n - p_E}{\sqrt{p_0(1-p_0)\left(\frac{2}{1M}\right)}} \quad (1)$$

This computation of Z statistics was performed in accordance with the SAGE (serial analysis of gene expression) data [60], where p_n and p_E are non-GFP and EXP7 RPKM values, respectively, divided by 1,000,000, and p_0 is the average of p_n and p_E .

Finally, a gene was identified as differentially expressed in EXP7 cells if *P* was less than 0.001 calculated by the Z statistic in each repeat and the EXP7/non-GFP fold change was greater than 2.

RNA-Seq data reported in the manuscript are available at the NCBI Sequence Read Archive (accession number [SRA045009.1]).

Detection of alternative-splicing events

Alternative-splicing events were computed using the RACKJ program [59]. A χ^2 value for goodness-of-fit was computed based on the reads supporting exon skipping (intron retention, alternative 5'/3' splice site) and the reads comprising one skipped exon (whole gene for intron retention, other alternative 5'/3' splice events) and a corresponding *P*-value was calculated. To classify enriched and reduced expression of specific features between two cell populations, the expression of each alternative-splicing event was normalized to the expression level of the gene in which the event occurred, using the equation:

$$(E_i/E_j)/(N_i/N_j) \quad (2)$$

for the *i*-th event of the *j*-th gene in EXP7 (E) and non-GFP (N) samples. Both significance (*P* < 0.05) and fold change (>2) were required to define a differential alternative-splicing event.

Primer design for the validation of alternative-splicing events

Splice junction-specific primers used to validate alternative-splicing events were designed to cover the predicted junction sites with the 3' terminus of the oligonucleotide spanning into the adjacent exon by about 5 to 10 nucleotides. Flanking primers were designed within the sequences of the closest adjacent exon if possible. The melting temperature of primers was selected in the range of 58 to 62°C. Primer pairs were selected using the Primer3 algorithm. The gene At4g26410 was used as reference for transcript normalization in accordance with our data and a previous study [61].

Quantitative RT-PCR

Total RNA was isolated (RNeasy Mini Kit; Qiagen) and treated with DNase (TURBO DNA-free Kit; Ambion Inc., Austin, TX, USA) in accordance with the manufacturer's instructions. cDNA was synthesized using DNA-free RNA with a commercial primer Oligo-dT(20); Invitrogen Corp., Carlsbad, CA, USA) and reverse transcriptase (Superscript II; Invitrogen Corp., Carlsbad, CA, USA). After incubation at 50°C for 1 hour and subsequently at 70°C for 15 minutes, 1 µl of RNase H was added and incubated for 20 minutes at 37°C. The cDNA

was used as a PCR template in a 20 μ l reaction system (SYBR Green PCR Master Mix; Applied Biosystems, Foster City, CA, USA) with programs recommended by the manufacturer in an automated sequencer (ABI Prism 7500 Sequence Detection System; Applied Biosystems). Three independent replicates were performed for each sample. The $\Delta\Delta C_t$ method was used to determine the relative amount of gene expression [62]. A list of all primers used in this study is provided (see Additional file 4, Table S7).

Gene Ontology analysis

Enrichment analysis of GO categories was performed with GOBU [15] using the TopGo 'elim' method [63] from the aspect 'biological process'. The elim algorithm iteratively removes from higher-level GO terms the genes mapped to significant terms, and thus keeps unimportant functional categories from being enriched.

Coexpression-based gene clustering

A set of 3,800 sample hybridizations using the ATH1 gene chip were collected from the NASCarrays database and normalized by the robust multiarray average (RMA) method using R software. A list of 56 genes with functions in root-hair development (see Additional file 4, Table S2) was used to identify expression signatures in the sample arrays that discriminate processes associated with root-hair development. We first identified in all hybridizations the 25% highest and the 25% lowest signals. We selected a subset of 111 hybridizations, in which more than 70% of the genes from the list were among those that gave the highest or lowest signals, and 635 differentially expressed genes were clustered based on the portion of these 111 hybridizations of the RMA data (see Additional file 4, Table S1). Coexpression relationships between genes were identified based on Pearson correlations of greater than or equal to 0.95. The software package is available with a thorough description [66].

Protein extraction, SDS-PAGE, and in-gel trypsin digestion

Protoplasts from EXP7 and non-GFP protoplasts were stored at -80°C , thawed on ice for 5 minutes, mixed by vortex several times, and suspended in 5 volumes of pre-cooled acetone (-20°C) containing 10% (v/v) trichloroacetic acid, and 0.07% (v/v) 2-mercaptoethanol. Proteins were then precipitated for 2 h at -20°C after thorough mixing. Proteins were collected by centrifuging at 35,000 g (JA-20 108 rotor; Beckman J2-HS; Beckman Coulter) at 4°C for 30 minutes. The supernatant was removed, and the protein pellets were washed three times with cold acetone containing 0.07% (v/v) 2-mercaptoethanol and 1 mmol/l phenylmethanesulfonyl fluoride. Protein pellets were dried by lyophilization and

stored at -80°C , or immediately extracted using Laemmli buffer (63 mM TrisHCl pH 6.8, 10% glycerol, 2% SDS, 0.0025% bromophenol blue). Protein concentration was determined using a protein assay kit (Pierce, Biotechnology Inc., Rockford, IL, USA).

Proteins from each sample were separated using a Bis-Tris precast gradient gel (4 to 6%, Bio-Rad Laboratories, Inc., Hercules, CA, USA) in accordance with the manufacturer's instructions. After electrophoresis, the gel was stained with Coomassie blue (Colloidal Blue Staining Kit; Invitrogen).

The protocol used for in-gel trypsin digestion of proteins in gels was adapted from the method described previously [64]. In brief, protein bands were manually excised, and each band was cut into small pieces (about 0.5 mm). The gel pieces were washed three times with a solution containing 50% methanol and 5% acetic acid for 2 hours, and washed twice with a solution containing 25 mmol/l NH_4HCO_3 in 50% acetonitrile for 10 minutes each, and then the gel pieces were dried in a vacuum centrifuge. Reduction of proteins in gel pieces was performed with dithiothreitol and alkylation with iodoacetamide, and the gel pieces were washed and dried in a vacuum centrifuge. A trypsin solution in 25 mmol/l NH_4HCO_3 containing 75 to 100 ng of sequencing-grade modified trypsin (Promega Corp., Madison, WI, USA) in 25 to 40 μ l of solution was added and incubated with the gel pieces for 12 to 16 hours at 37°C . To recover the tryptic peptides, a 30 μ l solution containing 5% formic acid and 50% acetonitrile was added to the gel pieces, agitated in a vortex for 30 to 60 min and withdrawn into a new tube. The recovery was repeated once with 15 μ l of solution, and the resulting two recoveries were combined and dried in a vacuum centrifuge. The dried pellet was redissolved in 10 to 20 μ l of 0.1% formic acid for LC-MS/MS analysis.

LC-MS/MS analysis

Liquid chromatography was performed (nanoACQUITY UPLC System; Waters Corp., Milford, MA, USA) coupled to a hybrid mass spectrometer (LTQ Orbitrap Velos; Thermo Fisher Scientific) equipped with a nanospray interface (PicoView; New Objective Inc., Woburn, MA, USA). Peptide mixtures were loaded onto a 75 $\mu\text{m} \times 250$ mm column packed with C18 resin (nanoACQUITY UPLC BEH130; Waters) and separated using a segmented gradient in 90 minutes using 5 to 40% solvent B (95% acetonitrile with 0.1% formic acid) at a flow rate of 300 nl/min. Solvent A was 0.1% formic acid in water. The samples were maintained at 4°C in the autosampler. The mass spectrometer (LTQ Orbitrap Velos; Thermo Fisher Scientific) was operated in the positive ion mode with the following acquisition cycle: a full scan (m/z 350 to 1600) recorded in the Orbitrap

analyzer at a resolution R of 60,000 was followed by MS/MS of the 20 most intense peptide ions in the LTQ analyzer.

Database search

Two search algorithms, Mascot (version 2.2.06; Matrix Science, Boston, MA, USA) and SEQUENT, which integrated in the Proteome Discoverer software (version 1.3.0.339; Thermo Fisher Scientific), were used to identify proteins. Searches were made against the Arabidopsis protein database annotated in TAIR10, concatenated with a decoy database containing the randomized sequences of the original database. Peak list data (MGF) files used for database searches were generated from Xcalibur raw files using a program in the MassMatrix conversion tools (version 1.4). The protein sequences in the database were searched with trypsin digestion at both ends, two missed cleavages allowed, fixed modifications of carbamidomethylation at Cys, variable modifications of oxidation at Met, methyl ester of the C-terminus, acetylation of both the N-terminus and lysine site, and phosphorylation at Ser, Thr, and Tyr. The merged MGF files were searched with 10 ppm of precursor peptide mass tolerance (monoisotopic) and 0.8 Da (Dalton) of MS/MS mass tolerance. The search results were passed through additional filters, with peptide confidence greater than 95%, before exporting the data. These filters resulted in a false discovery rate of less than 1% after decoy database searches were performed.

Quantification of proteins

Protein quantification was based on the NSAF method [30]. The NSAF for a protein P is calculated by the equation $P = k/L \times SpC$, where K is the number of spectral counts for one protein, L is the length of the protein, and SpC is the sum of all proteins in the experiment. For statistical analysis of the dataset, the natural log of each NSAF ratio of EXP7/non-GFP was calculated. Proteins with significant changes in abundance between two samples were selected using a method described by Cox and Mann [31]. The mean and standard deviation (SD) from the log₂ ratios of the 1,774 quantified proteins overlapping in both biological repeats was calculated. Next, 95% confidence intervals (CIs) (Z score = 1.96) were used to select proteins whose distribution was removed from the main distribution. For the downregulated proteins, the 95% CI was -0.2212 (mean ratio of the 1,774 proteins) $- 1.96 \times 1.2033$ (SD), corresponding to a protein ratio of 0.167. Similarly, for the upregulated proteins, the mean 95% CI was calculated (mean ratio $+ 1.96 \times$ SD), corresponding to a protein ratio of 4.399. Protein ratios outside this range were defined as being significantly different at $P < 0.05$. The threshold values were validated by power analysis with $\beta = 0.80$.

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [65] with the dataset identifier PXD000195 [29].

Additional material

Additional file 1: Figures S1-S8. Figure S1. Separation and size distribution of EXP7 protoplasts. Figure S2. Validation of RNA-sequencing (RNA-seq) results by quantitative reverse transcription (qRT)-PCR. Figure S3. Coexpression relationships of genes preferentially expressed in root hairs. Figure S4. Root-hair phenotypes of mutants harboring defects in the expression of genes from the coexpression network. Figure S5. Validation of alternative-splicing features by quantitative reverse transcription (qRT)-PCR. Figure S6. Scheme for the identification and quantification of proteins. Figure S7. Distributions of molecular weight and isoelectric point of proteins. Figure S8. The interactome of root hairs.

Additional file 2: Dataset S1. Read numbers and RPKM values from the transcriptome analysis.

Additional file 3: Dataset S2. Differentially expressed genes.

Additional file 4: Tables S1-S7

Additional file 5: Dataset S3. Splice junctions in EXP7 and non-GFP cells. GFP, green fluorescent protein.

Additional file 6: Dataset S4. Overview of alternative-splicing events in EXP7 and non-GFP cells. GFP, green fluorescent protein.

Additional file 7: Data set S5. Differential alternative-splicing events. Table S1. List of the root-related microarray experiments used for gene clustering. Table S2. Genes used for the search for expression signatures. Table S3. Post-translational modifications in the root protoplast proteome. Table S4. Root hair-specific proteins. Table S5. Differentially expressed proteins in EXP7 cells. Table S6. Confirmed interactions between root-hair proteins. Table S7. Primers used for the validation of alternatively splicing features and differentially expressed genes.

Abbreviations

baHLH: Basic helix-loop-helix; BSA: Bovine serum albumin; DADA: Differential alternative donor/acceptor; DAS: Differential alternative splicing; DES: Differential exon skipping; DIR: Differential intron retention; FACS: Fluorescence-activated cell sorting; GDP: Guanosine-5'-diphosphate; GFP: Green fluorescent protein; GO: Gene Ontology; GTP: Guanosine-5'-triphosphate; LC-MS/MS: Liquid chromatography/tandem mass spectrometry; MES: 2-(N-morpholino)ethanesulfonic acid; NCBI: National Center for Biotechnology Information; NMD: Nonsense-mediated decay; NSAF: Normalized spectral abundance factor; qRT: Quantitative reverse transcription; PPI: Protein-protein interaction; psi: Pounds per square inch; PTM: post-translational modification; RMA: Robust multiarray average; ROP: Rho-type; RPKM: Reads per 1 Kbps of exon model per million mapped reads; RUST: Regulated unproductive splicing and translation; SAGE: Serial analysis of gene expression; SJ: splice junction; SpC: spectral counts; TCA: Tricarboxylic acid

Authors' contributions

PL performed the proteomic analysis including data analysis. WL optimized the protocol for protoplasting and cell sorting, prepared the RNA samples, and validated the RNA-seq data by qRT-PCR. W-DL developed the software used for the calculation of RPKM values and splicing patterns and processed the raw data. WL, PL, and WS analyzed the RNA-seq data. WS and SS conceived of and supervised the study. WS wrote the manuscript with input from PL, WL, and SS. All authors read and approved the final manuscript.

Acknowledgements

Proteomic mass spectrometry analyses were performed by the Core Facilities for Proteomics Research located at the Institute of Biological Chemistry, Academia Sinica, supported by a National Science Council grant (NSC 98-

3112-B-001-023, NSC 99-3112-B-001-025) and Academia Sinica. In-gel digestion was performed with the help of the Proteomics Core Laboratory sponsored by the Institute of Plant and Microbial Biology and the Agricultural Biotechnology Research Center, Academia Sinica. Cell sorting was performed by Su-Hsin Huang at the Scientific Instrument Center's Flow Cytometry Core Facility in the Institute of Plant and Microbial Biology, Academia Sinica. We thank H. Rouached and F. Gosti (Supagro Montpellier, France) for discussion, T. J. Buckhout (Humboldt University Berlin, Germany) for critical comments on the manuscript, and A. Kuo (Schmidt laboratory) for figure artwork and manuscript editing. Expert technical help was provided by Y. Y. Liao, J. Y. Hsiao, C. Cheng (Schmidt laboratory) and G. Fu (Bioinformatics Core Facility in the Institute of Plant and Microbial Biology, Academia Sinica). This work was supported by an Academia Sinica thematic grant (number 02234g) to WS and SS.

Authors' details

¹Institute of Plant and Microbial Biology, Academia Sinica, 11529 Taipei, Taiwan. ²Dipartimento di Scienze Agrarie e Ambientali, Università degli Studi di Udine, Via delle Scienze 206, I-33100 Udine, Italy. ³Biotechnology Center, National Chung-Hsing University, Taichung, Taiwan. ⁴Genome and Systems Biology Degree Program, College of Life Science, National Taiwan University, Taipei, Taiwan.

Received: 7 February 2013 Revised: 29 March 2013

Accepted: 25 June 2013 Published: 25 June 2013

References

- Dolan L: Positional information and mobile transcriptional regulators determine cell pattern in the Arabidopsis root epidermis. *J Exp Bot* 2006, **57**:51-54.
- Wada T, Tachibana T, Shimura Y, Okada K: Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC. *Science* 1997, **277**:1113-1116.
- Lee MM, Schiefelbein J: WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. *Cell* 1999, **99**:473-483.
- Grebe M: The patterning of epidermal cells in Arabidopsis-updated. *Curr Opin Plant Biol* 2012, **15**:31-37.
- Grierson C, Schiefelbein J: Root hairs. *Arabidopsis Book* 2002, **1**:e0060.
- Ishida T, Kurata T, Okada K, Wada T: A genetic regulatory network in the development of trichomes and root hairs. *Annu Rev Plant Biol* 2008, **59**:365-386.
- Jones M, Smirnov N: Nuclear dynamics during the simultaneous and sustained tip growth of multiple root hairs arising from a single root epidermal cell. *J Exp Bot* 2006, **57**:4269-4275.
- Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN: A gene expression map of the Arabidopsis root. *Science* 2003, **302**:1956-1960.
- Brady SM, Orlando DA, Lee JY, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN: A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 2007, **318**:801-806.
- Gifford ML, Dean A, Gutierrez RA, Coruzzi GM, Birnbaum KD: Cell-specific nitrogen responses mediate developmental plasticity. *Proc Natl Acad Sci USA* 2008, **105**:803-808.
- Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN: Cell identity mediates the response of Arabidopsis roots to abiotic stress. *Science* 2008, **320**:942-945.
- Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, Zheng X, Wang JY, Lee MM, Benfey P, Woolf PJ, Schiefelbein J: A gene regulatory network for root epidermis cell differentiation in Arabidopsis. *PLoS Genet* 2012, **8**:e1002446.
- Petricka JJ, Schauer MA, Megraw M, Breakfield NW, Thompson JW, Georgiev S, Soderblom EJ, Ohler U, Moseley MA, Grossniklaus U, Benfey PN: The protein expression landscape of the Arabidopsis root. *Proc Natl Acad Sci USA* 2012, **109**:6811-6818.
- Cho HT, Cosgrove DJ: Regulation of root hair initiation and expansin gene expression in Arabidopsis. *Plant Cell* 2002, **14**:3237-3253.
- Lin Y, Ulanov AV, Lozovaya V, Widholm J, Zhang G, Guo J, Goodman HM: Genetic and transgenic perturbations of carbon reserve production in Arabidopsis seeds reveal metabolic interactions of biochemical pathways. *Planta* 2006, **225**:153-164.
- Masucci JD, Schiefelbein JW: The *rhd6* Mutation of *Arabidopsis thaliana* alters root-hair initiation through an auxin- and ethylene-associated process. *Plant Physiol* 1994, **106**:1335-1346.
- Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, Linstead P, Schaefer DG, Dolan L: An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 2007, **316**:1477-1480.
- Carol RJ, Takeda S, Linstead P, Durrant MC, Kakesova H, Derbyshire P, Drea S, Zarsky V, Dolan L: A RhoGDP dissociation inhibitor spatially regulates growth in root hair cells. *Nature* 2005, **438**:1013-1016.
- Lin WD, Liao YY, Yang TJ, Pan CY, Buckhout TJ, Schmidt W: Coexpression-based clustering of Arabidopsis root genes predicts functional modules in early phosphate deficiency signaling. *Plant Physiol* 2011, **155**:1383-1402.
- Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S: NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res* 2004, **32**:D575-577.
- Baumberger N, Ringli C, Keller B: The chimeric leucine-rich repeat/extensin cell wall protein LRX1 is required for root hair morphogenesis in *Arabidopsis thaliana*. *Genes Dev* 2001, **15**:1128-1139.
- Grierson CS, Roberts K, Feldmann KA, Dolan L: The COW1 locus of Arabidopsis acts after RHD2, and in parallel with RHD3 and TIP1, to determine the shape, rate of elongation, and number of root hairs produced from each site of hair formation. *Plant Physiol* 1997, **115**:981-990.
- Santi S, Schmidt W: Dissecting iron deficiency-induced proton extrusion in Arabidopsis roots. *New Phytol* 2009, **183**:1072-1084.
- Kim DW, Lee SH, Choi SB, Won SK, Heo YK, Cho M, Park YI, Cho HT: Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns. *Plant Cell* 2006, **18**:2958-2970.
- Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC: Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res* 2010, **20**:45-58.
- Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M: Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res* 2012, **22**:1184-1195.
- Reddy AS: Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* 2007, **58**:267-294.
- Carvalho RF, Feijao CV, Duque P: On the physiological significance of alternative splicing events in higher plants. *Protoplasma* 2012.
- Data set S6., Proteins and peptides identified in EXP7 and non-GFP cells (available at ProteomeXchange Consortium; dataset identifier PXD000195).
- Zybailov B, Mosley AL, Sardu ME, Coleman MK, Florens L, Washburn MP: Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* 2006, **5**:2339-2347.
- Cox J, Mann M: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008, **26**:1367-1372.
- Arabidopsis Interactome Mapping. C: Evidence for network evolution in an Arabidopsis interactome map. *Science* 2011, **333**:601-607.
- Lan P, Li W, Schmidt W: Complementary proteome and transcriptome profiling in phosphate-deficient Arabidopsis roots reveals multiple levels of gene regulation. *Mol Cell Proteomics* 2012, **11**:1156-1166.
- Hsieh SI, Castruita M, Malasarn D, Urzica E, Erde J, Page MD, Yamasaki H, Casero D, Pellegrini M, Merchant SS, Loo JA: The proteome of copper, iron, zinc, and manganese micronutrient deficiency in *Chlamydomonas reinhardtii*. *Mol Cell Proteomics* 2013, **12**:65-86.
- Jones MA, Raymond MJ, Smirnov N: Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in Arabidopsis. *Plant J* 2006, **45**:83-100.
- Yi K, Menand B, Bell E, Dolan L: A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. *Nat Genet* 2010, **42**:264-267.
- Fu Y, Gu Y, Zheng Z, Wasteneys G, Yang Z: Arabidopsis interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. *Cell* 2005, **120**:687-700.
- Oyama T, Shimura Y, Okada K: The IRE gene encodes a protein kinase homologue and modulates root hair growth in Arabidopsis. *Plant J* 2002, **30**:289-299.
- Müller M, Schmidt W: Environmentally induced plasticity of root hair development in Arabidopsis. *Plant Physiol* 2004, **134**:409-419.

40. Monshausen GB, Bibikova TN, Messerli MA, Shi C, Gilroy S: **Oscillations in extracellular pH and reactive oxygen species modulate tip growth of Arabidopsis root hairs.** *Proc Natl Acad Sci USA* 2007, **104**:20996-21001.
41. Messerli MA, Danuser G, Robinson KR: **Pulsatile influxes of H⁺, K⁺ and Ca²⁺ + lag growth pulses of *Lilium longiflorum* pollen tubes.** *J Cell Sci* 1999, **112**(Pt 10):1497-1509.
42. Gillis J, Pavlidis P: **"Guilt by association" is the exception rather than the rule in gene networks.** *PLoS Comput Biol* 2012, **8**:e1002444.
43. Shin R, Berg RH, Schachtman DP: **Reactive oxygen species and root hairs in Arabidopsis root response to nitrogen, phosphorus and potassium deficiency.** *Plant Cell Physiol* 2005, **46**:1350-1357.
44. Deal RB, Henikoff S: **A simple method for gene expression and chromatin profiling of individual cell types within a tissue.** *Dev Cell* 2010, **18**:1030-1040.
45. Grotz N, Fox T, Connolly E, Park W, Guerinot ML, Eide D: **Identification of a family of zinc transporter genes from Arabidopsis that respond to zinc deficiency.** *Proc Natl Acad Sci USA* 1998, **95**:7220-7224.
46. Hodges ME, Wickstead B, Gull K, Langdale JA: **Conservation of ciliary proteins in plants with no cilia.** *BMC Plant Biol* 2011, **11**:185.
47. Lewis BP, Green RE, Brenner SE: **Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans.** *Proc Natl Acad Sci USA* 2003, **100**:189-192.
48. Lareau LF, Brooks AN, Soergel DA, Meng Q, Brenner SE: **The coupling of alternative splicing and nonsense-mediated mRNA decay.** *Adv Exp Med Biol* 2007, **623**:190-211.
49. Wang BB, Brendel V: **Molecular characterization and phylogeny of U2AF35 homologs in plants.** *Plant Physiol* 2006, **140**:624-636.
50. Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, Marshall J, Fuller J, Cardle L, McNicol J, Dinh HQ, Barta A, Brown JW: **Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis.** *Nucleic Acids Res* 2012, **40**:2454-2469.
51. Cheung AY, Niroomand S, Zou Y, Wu HM: **A transmembrane formin nucleates subapical actin assembly and controls tip-focused growth in pollen tubes.** *Proc Natl Acad Sci USA* 2010, **107**:16390-16395.
52. Yi K, Guo C, Chen D, Zhao B, Yang B, Ren H: **Cloning and functional characterization of a formin-like protein (AtFH8) from Arabidopsis.** *Plant Physiol* 2005, **138**:1071-1082.
53. Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S: **Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics.** *Science* 2008, **320**:938-941.
54. Wang Q, Zhang Y, Yang C, Xiong H, Lin Y, Yao J, Li H, Xie L, Zhao W, Yao Y, Ning ZB, Zeng R, Xiong Y, Guan KL, Zhao S, Zhao GP: **Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux.** *Science* 2010, **327**:1004-1007.
55. Yang XJ, Seto E: **Lysine acetylation: codified crosstalk with other posttranslational modifications.** *Mol Cell* 2008, **31**:449-461.
56. Zhao S, Xu W, Jiang W, Yu W, Lin Y, Zhang T, Yao J, Zhou L, Zeng Y, Li H, Li Y, Shi J, An W, Hancock SM, He F, Qin L, Chin J, Yang P, Chen X, Lei Q, Xiong Y, Guan KL: **Regulation of cellular metabolism by protein lysine acetylation.** *Science* 2010, **327**:1000-1004.
57. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**:621-628.
58. Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12**:656-664.
59. Li W, Lin W-D, Ray P, Lan P, Schmidt W: **Genome-wide detection of condition-sensitive alternative splicing in Arabidopsis roots.** *Plant Physiol*, DOI:10.1104/pp.113.217778.
60. Kal AJ, van Zonneveld AJ, Benes V, van den Berg M, Koerkamp MG, Albermann K, Strack N, Ruijter JM, Richter A, Dujon B, Ansoorge W, Tabak HF: **Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources.** *Mol Biol Cell* 1999, **10**:1859-1872.
61. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR: **Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis.** *Plant Physiol* 2005, **139**:5-17.
62. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method.** *Methods* 2001, **25**:402-408.
63. Alexa A, Rahnenfuhrer J, Lengauer T: **Improved scoring of functional groups from gene expression data by decorrelating GO graph structure.** *Bioinformatics* 2006, **22**:1600-1607.
64. Shevchenko A, Wilm M, Vorm O, Jensen ON, Podtelejnikov AV, Neubauer G, Shevchenko A, Mortensen P, Mann M: **A strategy for identifying gel-separated proteins in sequence databases by MS alone.** *Biochem Soc Trans* 1996, **24**:893-896.
65. Vizcaino JA, Côté RG, Csordas A, Dianas JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O'Kelly G, Schoenegger A, Ovelleiro D, Pérez-Riverol Y, Reisinger F, Ríos D, Wang R, Hermjakob H: **The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013.** *Nucleic Acids Res* 2013, **41**:D1063-D1069.
66. Lin W, Liao YY, Yang JW, Pan CY, Buckhout T, Schmidt W: **Coexpression-based Clustering of Arabidopsis Root Genes Predicts Functional Modules in Early Phosphate Deficiency Signaling.** *Plant Physiol* 2011, **155**:1383-1402.

doi:10.1186/gb-2013-14-6-r67

Cite this article as: Lan et al.: Mapping gene activity of Arabidopsis root hairs. *Genome Biology* 2013 **14**:R67.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

