

# A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies

Yuling Jiao<sup>1,3,4</sup>, S Lori Tausta<sup>1,4</sup>, Neeru Gandotra<sup>1,4</sup>, Ning Sun<sup>2,4</sup>, Tie Liu<sup>1,3,4</sup>, Nicole K Clay<sup>1,3</sup>, Teresa Ceserani<sup>1,3</sup>, Meiqin Chen<sup>1,3</sup>, Ligeng Ma<sup>1,3</sup>, Matthew Holford<sup>2</sup>, Hui-yong Zhang<sup>1,3</sup>, Hongyu Zhao<sup>2</sup>, King-Wang Deng<sup>1</sup> & Timothy Nelson<sup>1</sup>

**The functions of the plant body rely on interactions among distinct and nonequivalent cell types. The comparison of transcriptomes from different cell types should expose the transcriptional networks that underlie cellular attributes and contributions. Using laser microdissection and microarray profiling, we have produced a cell type transcriptome atlas that includes 40 cell types from rice (*Oryza sativa*) shoot, root and germinating seed at several developmental stages, providing patterns of cell specificity for individual genes and gene classes. Cell type comparisons uncovered previously unrecognized properties, including cell-specific promoter motifs and coexpressed cognate binding factor candidates, interaction partner candidates and hormone response centers. We inferred developmental regulatory hierarchies of gene expression in specific cell types by comparison of several stages within root, shoot and embryo.**

The systems perspective emerging in biology promises to explain many behaviors of cells and organisms based on modular networks of expression, interaction, regulation and metabolism<sup>1</sup>. A study of *Arabidopsis thaliana* root cell-specific transcriptomes recently demonstrated the value of rigorously comparable cellular data sets for identifying the genes and networks responsible for root cell-specific functions and interactions<sup>2,3</sup>. Several other plant studies have compared transcriptomes to resolve transcripts specific for stages or locations<sup>4</sup>, but all are limited in scope to a few cell types or encompass zones or tissues combining multiple cell types.

We obtained transcriptome data for 40 rice cell types, comprising most types in seedling roots and shoots and in germinating seed. We acquired cell types by laser microdissection. Because laser microdissection does not require cell-specific tags or unique genetic lines, we were able to select cells entirely on the basis of their appearance and

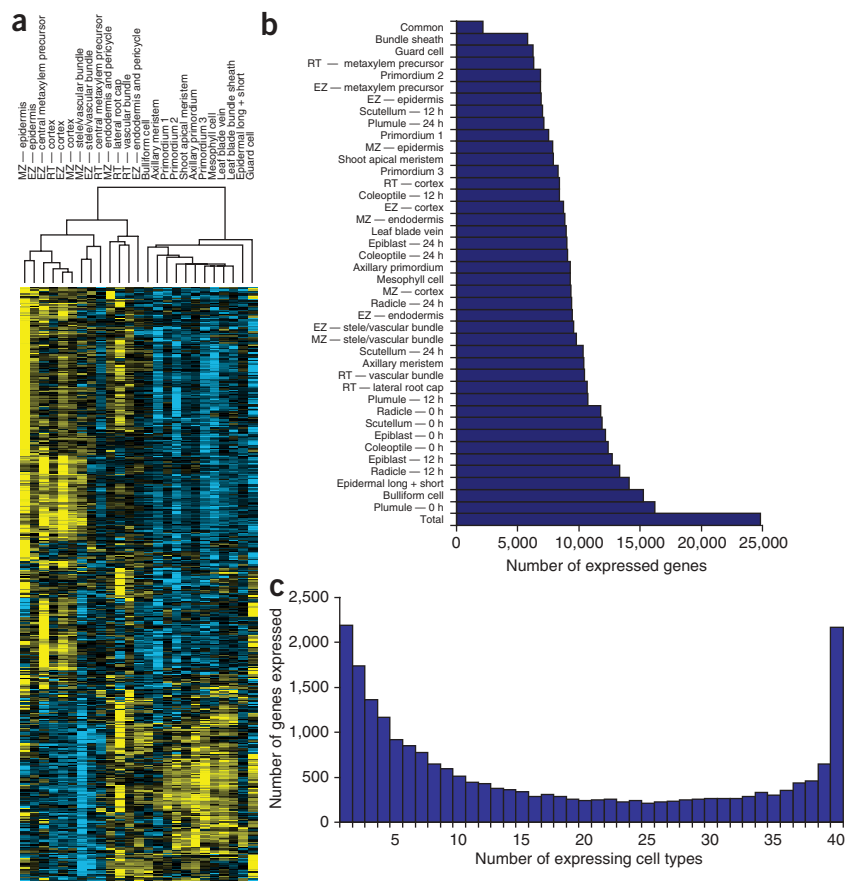
location in histological sections<sup>5</sup>. Among the 30,731 genes represented on the whole-genome long-oligonucleotide microarray platform, we detected the expression of 24,822 genes (80.8%) in at least one of the 40 cell types. A reference RNA pool provided a second probe against each cell type, allowing normalization of the entire database for quantitative comparisons between any or all cell types. We developed statistical procedures to address two atlas-related issues: (i) the transcript abundance for most expressed genes differed significantly between each experimental cell type and the control reference RNA to which it was compared, and (ii) each cell type-expressed gene set overlapped but was not identical to the common reference gene set. Cell types are described in **Supplementary Table 1** online (see also URLs section of Methods).

The atlas includes cell types from shoots, roots and germinated seeds, including three time points from seed imbibition (0 h, 12 h and 24 h), three root developmental zones (tip, elongation zone and maturation zone) and several leaf stages (leaf primordia P1, P2, P3 and P5 from 5- to 7-d-old seedlings). We validated the transcriptional profiles by RT-PCR, using an aliquot of the laser microdissection-isolated RNA to compare expression of representative genes (**Supplementary Table 2** online). The microarray data were consistent with published expression patterns of specific rice genes and of *A. thaliana* orthologs (for example, thioredoxin-h<sup>6</sup>).

Cell-specific transcriptomes showed qualitative and quantitative differences consistent with functional specialization. The transcriptome of each cell type was distinct, consisting of transcripts from 6,000–16,000 genes (26%–52% of all genes represented on the array, **Fig. 1a**). Many genes that were undetectable in the transcriptome of an entire organ were present at substantial amounts in one or more cell types within that organ. For example, we identified 879 genes, including several genes encoding transcription factors, that were expressed in one or more leaf cell types but that were below the

<sup>1</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA, and Peking-Yale Joint Research Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China. <sup>2</sup>Center for Statistical Genomics and Proteomics, Yale University, New Haven, Connecticut 06520, USA. <sup>3</sup>Present addresses: Division of Biology, California Institute of Technology, Pasadena, California 91125, USA (Y.J.); Biology Department, Stanford University, Stanford, California 94305, USA (T.L.); Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA (N.K.C.); Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada (T.C.); Peking-Yale Joint Research Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China (M.C.); National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing 102206, China (L.M. and H.-Y.Z.). <sup>4</sup>These authors contributed equally to this work. Correspondence should be addressed to T.N. (timothy.nelson@yale.edu).

Received 5 August 2008; accepted 17 October 2008; published online 4 January 2009; doi:10.1038/ng.282



**Figure 1** Global patterns of cellular gene expression. **(a)** Cluster display of differentially expressed genes in 40 cell types. Complete linkage hierarchical clustering was performed on relative expression (z scores) as described in Methods. Only gene models that showed differential expression in at least one cell type were included. Yellow, high expression in a specific cell type; blue, low expression. The dendrogram shows the relationship among cell types based on expression. RT, root tip; EZ, elongation zone; MZ, maturation zone; epidermal long + short: epidermal pavement cells minus guard and bulliform cells. **(b)** Number of genes expressed in each of the 40 cell types. Total: number expressed in at least one cell type; common: genes expressed in all 40 cell types. **(c)** Number of genes expressed in multiple cell types.

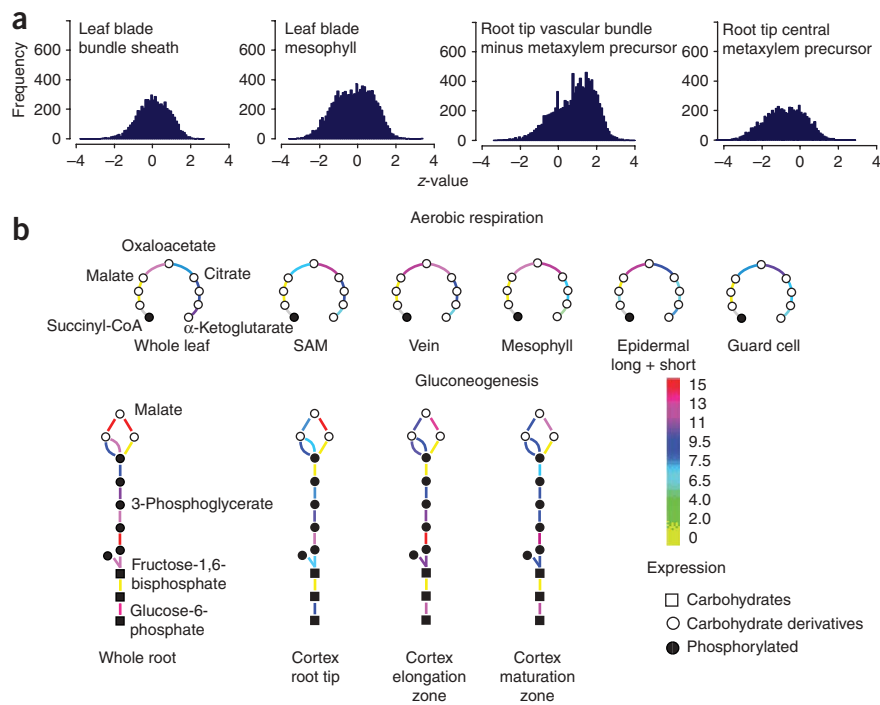
(52%) in bulliform cells, with a median of 8,831 (29% of all genes expressed in the 40 cells). The transcriptomes of regions of germinating seeds included a larger number of genes, with a median of 10,733 (35%), reflecting the broad range of activities associated with energy mobilization and the rapid building of all seedling systems (**Fig. 1b**). We found that 2,188 genes (7% of all genes) were expressed in only a single cell type among the 40 measured, and 2,171 genes (7%) were expressed in all 40 (**Fig. 1c**).

The sets of expressed and unexpressed genes were highly biased in certain cell

threshold of detection in RNA from whole leaves (**Supplementary Fig. 1** online). An organ-tissue-cell hierarchical relationship among cell type transcriptomes was evident (**Fig. 1a**). Transcriptomes of different cells within shoot, root or embryo were more similar to each other than to those from other tissues. Seedling cell types expressed fewer genes, ranging from <6,000 (20%) in leaf bundle sheath cells to ~16,000

types, including many transcripts that were cell specific or nearly cell-specific in their presence or absence (**Fig. 2a** and **Supplementary Fig. 2** online). Examples of cells enriched in transcripts with high cell

**Figure 2** Cell-specific transcripts and selected metabolic pathways. **(a)** Histograms of over- and underrepresented transcripts ( $z > 0$  and  $z < 0$ , respectively), relative to the average of all cell types. Bundle sheath and mesophyll cells have normal distributions; root tip vascular bundle (skewed positive) and metaxylem precursor (skewed negative) are enriched in transcripts that are cell specific in their presence or absence from the average transcriptome. Criteria for z plot are described in Methods. Plots for all 40 cell types are shown in **Supplementary Figure 2**. **(b)** Transcript abundance for steps in respiration and gluconeogenesis in several cell types. Pathways are depicted as nodes for metabolites and colored lines for enzyme transcript abundance. Expression-level change of each reaction is shown in a color relative to the expression level. Gray lines represent missing gene annotation or expression data. For such housekeeping genes, expression is ubiquitous and relatively uniform in all cell types.



specificity ( $z$  values, see **Supplementary Methods** online) were root tip vascular bundle (1,472 genes, or 14% of the genes detected in that cell), root maturation zone epidermis (1,017 or 13%), bulliform cells (389 or 2.5%) and plumule (0 h) (699 or 4.3%). Cell types with restricted roles (such as root metaxylem precursor or long and short epidermal cell) expressed a narrower and more exclusive set of genes

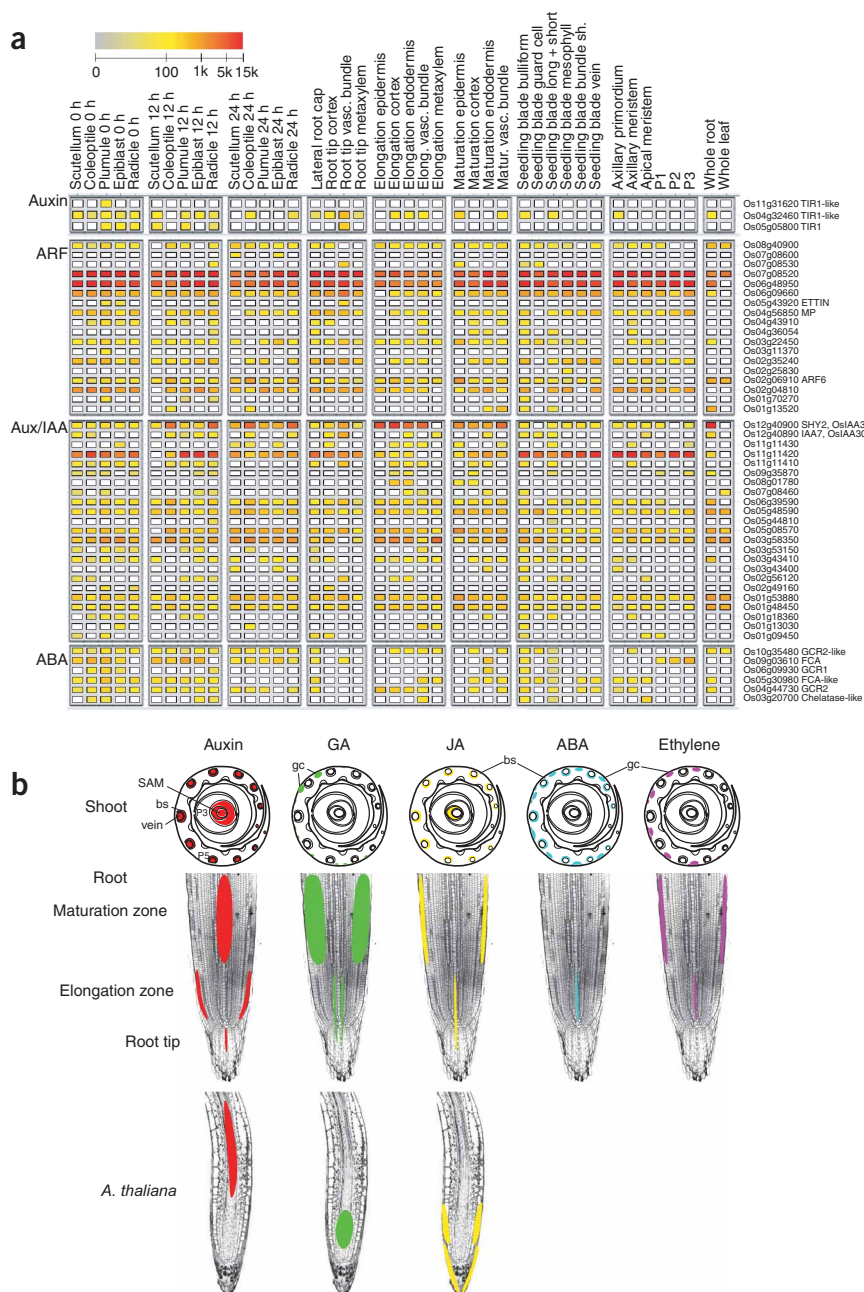
than did other types (for example, leaf mesophyll or shoot apical meristem, **Fig. 2a**). Highly specific gene sets corresponded well with known cell functions (for example, CVP2-like and nucleolins in metaxylem precursor<sup>7,8</sup>).

Rice cell transcriptomes showed frequent chromosomal clustering of two to four coexpressed genes, as has been documented in *A. thaliana*<sup>9</sup> and other eukaryotes<sup>10</sup>. The set of cell-specifically expressed genes showed this pattern, although without a clustering discernibly related to cell function (**Supplementary Table 3** online).

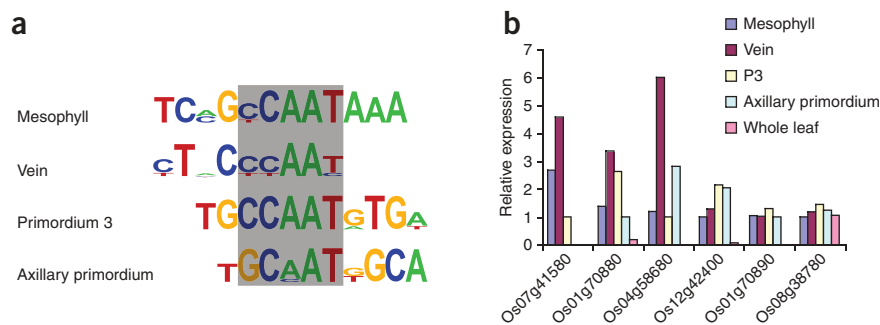
Several cell types had transcriptomes enriched in Gene Ontology (GO) classes reflecting known cellular functions (**Supplementary Fig. 3** online). For example, photosynthesis-related GO terms were enriched in leaf mesophyll and bundle sheath cells, and shoot morphogenesis GO terms were associated with cells of shoot meristem and primordia. Some enrichments suggested localized activities that were previously uncharacterized. For instance, protein translation-related GO terms were enriched in root elongation zone cortex and central metaxylem precursor cells.

We identified candidate housekeeping genes statistically based on ubiquitous and uniform expression across all cell types, a successful criterion for identifying housekeeping genes in other organisms<sup>11</sup>. Depending on the window chosen to filter uniformity of expression, the housekeeping set included 248–1,588 genes (1%–6% of all expressed genes) (see URLs section of Methods). Within this set, which included most expected housekeeping functions (for example, ubiquitin and heat shock proteins), the genes that encoded steps of basic metabolic pathways, such as gluconeogenesis and respiration, were notable in their uniformity of expression among all cell types (**Fig. 2b**). The housekeeping set was enriched in hormone-related genes for six out of seven phytohormones, probably reflecting the fundamental roles of phytohormones in the viability and growth of all plant cells. Auxin-related genes were not a significant part of the ubiquitously expressed set, evidence of the spatial and temporal focus of auxin actions.

The comparison of a developmental series of site-specific transcriptomes provided a wealth of genes with candidate developmental roles. For example, for leaf primordia P1, P2 and P3, 1,129 genes were expressed in all three stages at relatively constant amounts ( $\leq 1.5$ -fold change), whereas 337, 882 and 848 genes showed reproducible maxima in P1, P2 or P3, and 22 and 76 genes showed P1-P2-P3 downward or upward transitions, respectively, of more than twofold between steps (**Supplementary Fig. 4** online). These classes included



**Figure 3** Cellular distributions of transcripts from selected hormone-related genes. **(a)** Heat map representation of abundance of transcripts from auxin-related and ABA-related genes in 40 cell types. Heat maps related to other hormones are shown in **Supplementary Figure 6**. Several genes are identified by *A. thaliana* ortholog. **(b)** Diagrams of centers of expression for hormone-related genes. “Shoot” is a transverse section through shoot apical meristem (SAM), leaf primordia (P2, P3) and seedling leaves showing vascular bundles. Gc, guard cell; bs, bundle sheath. “Root” shows cellular locations of enriched hormone related transcripts superimposed over longitudinal images of root developmental zones. This figure summarizes hormone-related gene data in **Supplementary Table 4**. Data for the *A. thaliana* summary are from ref. 2.



**Figure 4** Identification of candidate *cis* and *trans* transcriptional control cognate partners on the basis of cellular coexpression. **(a)** CCAAT box variants and **(b)** NF-Y family members enriched in different leaf cell types. Sequence logos in **a** represent a motif matrix for each cell type, as described in Methods. The commonly shared CCAAT core motif is highlighted in the gray box.

transcription factors, F-box proteins, RNA-binding proteins and others with potentially regulatory roles.

Similarly, within root developmental zones, cell types showed sets of genes with coordinate developmental transitions (**Supplementary Fig. 5** online). For example, in root vascular cells, we identified 440 and 128 genes that decreased or increased, respectively, by more than twofold in transcript abundance from tip to elongation zone to maturation zone. In root cortex cells, transcripts for 75 genes decreased and 52 increased in the same zones. The genes with these developmental profiles overlapped substantially with those found in a developmental study of *A. thaliana* root cell types<sup>3</sup>, for cases in which gene orthologs could be matched. For example, 77/112 identified *A. thaliana* orthologs of a rice gene set (1,126 genes) that showed more than twofold higher expression in vascular bundle than in cortex satisfied the same criteria in the *A. thaliana* data sets (see URLs section of Methods). This correspondence of ortholog expression patterns is notable considering the evolutionary distance between the monocot rice and the dicot *Arabidopsis*.

Transcription from hormone-responsive and hormone receptor genes was enriched in specific cell types and organ zones. We examined the sets of genes related to seven phytohormones in rice, identified by their up- or downregulation by hormone treatments of rice cell cultures (refs. 12,13 and the Rice Expression Database), or by their annotation (as reported on The Institute for Genomic Research website; see URLs section in Methods). Genes in these classes showed cell type patterns of enrichment, including ‘activity centers’ in shoots and roots (**Fig. 3** and **Supplementary Fig. 6** and **Supplementary Table 4** online). Cells whose transcriptomes were particularly hormone-biased were guard cells (giberellic acid- and abscisic acid (ABA)-related), bundle sheath (ABA), metaxylem precursor (ABA), shoot apical meristem (auxin) and leaf primordium (auxin). As noted above, hormone-regulated genes, both induced and repressed, were substantially enriched within the set of genes expressed in all 40 cell types.

The cellular and spatial patterns of hormone-related transcripts in rice roots were similar to but distinct from the patterns reported for *A. thaliana* roots<sup>2</sup> (**Fig. 3b**). In rice, an auxin activity center was evident in maturation zone vascular bundle cells and endodermis, similar to the stage 3 stele center in *A. thaliana*. An apparent giberellic acid activity center in rice elongation zone vascular bundle cells and metaxylem precursor cells is similar to a center in *A. thaliana* stage 2 stele. The rice giberellic acid center extends down to root tip vascular bundle cells and up to maturation zone epidermis and cortex, whereas the *A. thaliana* giberellic acid center extends to elongation zone

(stage 2) endodermis and cortex. Centers for jasmonic acid-related transcripts differed in rice and *A. thaliana*. In rice, the jasmonic acid center spans the epidermal cells of the maturation zone and metaxylem precursor cells in both the elongation zone and root tip. In *A. thaliana*, the jasmonic acid center is earlier (stages 1 and 2: elongation zone and tip) but also in the epidermis (and lateral root cap). Centers for ABA- and ethylene-related transcripts were evident in rice roots but not in *A. thaliana*.

Many hormone receptors show cell type-biased patterns of transcripts (**Fig. 3** and **Supplementary Fig. 6**). For example, three ABA receptors have been characterized in *A. thaliana*: the intracellular FCA and Mg-

chelatase subunit H and the plasma membrane-localized GCR2<sup>14–16</sup>. In rice, we found that putative homologs of FCA (Os05g30980) and GCR2 (Os10g35480) were highly expressed in guard cells. Distinct GCR2-like genes (Os10g35480 and Os04g44730) were expressed in root cell types where ABA is made and transported through the xylem under water stress conditions. In contrast, the rice chelatase-like homolog (Os03g20700) was expressed in apical meristem and elongation vascular bundle (**Fig. 3**). A distinct cell type distribution was evident for each member of the brassinosteroid receptor family (**Supplementary Fig. 6a**).

Cell-specific transcriptomes uncover genes with specific cellular patterns of expression, multigene expression fingerprints for particular cell types, genes coexpressed in a cell type (for example, to validate interaction partners or pathway candidates) and DNA motifs associated with cellular patterns of expression. As a demonstration, we searched for motifs enriched in the 1-kb upstream region of genes specifically expressed or enriched in particular cell types. Candidate regulatory sequence motifs of known or unknown sequence were associated with cell-specific transcripts in 18 out of 25 cell types examined. Among these were CCAAT box elements for leaf sets and dehydration-responsive and pathogen-responsive elements for root sets<sup>17</sup>, suggesting that stress and pathogens act on some motifs with basic cellular roles. Two dehydration-responsive elements (with MYB and MYC recognition sites) previously associated functionally with ABA-induced gene expression<sup>17</sup> were identified statistically from our data on the basis of their enrichment in transcriptomes of root vascular cells, a putative ABA-responsive center, and in transcriptomes of cortex cells, which also showed enrichment of ABA-responsive gene transcripts.

The CCAAT box, a common *cis*-acting element in promoters and enhancer regions<sup>18</sup>, showed a marked pattern of occurrence in transcripts specific for several shoot cell types. The CCAAT box and its binding proteins have complex roles in regulating plant gene transcription during plant development<sup>19</sup>. We found that the CCAAT box core motif present in many genes that are coexpressed in specific leaf cell types was flanked by a sequence that was specific for each leaf cell type, suggesting it has functional significance (**Fig. 4a**).

Cell-specific coexpression may prove to be a substantial aid in sorting true protein-DNA and protein-protein interaction partners from numerous candidates. Regulation of the CCAAT box is achieved by the binding of the highly conserved heterotrimeric nuclear factor Y (NF-Y) complex<sup>20</sup>. In *A. thaliana*, 36 NF-Y family members have been identified, most of which are expressed tissue specifically<sup>21</sup>. In rice, 28 NF-Y subunits have been annotated. Many rice cell type transcriptomes showed enrichment of a specific CCAAT box variant along with

certain NF-Y subunit variants, compared to total leaf, making these likely cognate pairs (Fig. 4). For instance, NF-Y variants Os04g58680 and Os07g41580 were enriched in leaf vein, where transcripts from genes with the potential target motif CTACCCAAT were enriched.

In summary, we provide numerous examples from rice in which an atlas of rigorously comparable cell-specific transcriptomes makes possible new views of the properties of cell types, including patterns of transcription and regulation not evident in data obtained from whole organs or tissues. A cell transcriptome atlas can aid in the formulation and validation of interaction and pathway networks by providing evidence of the coexpression of potential pathway members within the same cell type. Because plants consist of relatively few cell types, it should be possible to extend cell type atlases to include rigorously comparable molecular profiles of all cell types from plants grown under a variety of conditions and developmental states. As data sets for more cells and conditions are added, it should be possible to distinguish an invariant expression fingerprint for each cell type that will yield insight into the nature of cell types themselves.

## METHODS

**Plant growth and histological preparation.** Rice cell cultures for preparation of common reference RNA aliquots were grown as described previously<sup>22</sup>. Surface-sterilized rice seeds (*Oryza sativa* L. ssp. *japonica*) were incubated in darkness at 37 °C for 2 d in sterile distilled water and then transferred to MS medium for growth in a short-day cycle of 10 h light (120  $\mu\text{E m}^{-2} \text{s}^{-1}$ ) at 30 °C and 14 h darkness at 24 °C. For tissue fixation, germinated seed stages (0 h, 12 h and 24 h after imbibition) were processed whole. Roots, second true leaves (P5) and shoot apical meristem regions, including the leaf primordia P1, P2 and P3, were dissected from 5-d-old seedlings (counting from time of transfer to light) into samples  $\leq 4$  mm in thickness. Samples were processed to 8- $\mu\text{m}$  sections as described in **Supplementary Methods**. Before cell microdissections, we routinely evaluated RNA isolated from scrapes of test slides prepared in the same batch to assure that it was intact, using RNA methods described below. For all tissue samples, source tissues were archived in paraffin blocks stored at 4 °C to permit validation studies or subsequent cell harvests from the same material.

**Laser microdissection.** Cells were isolated by microdissection with Veritas and Pix-Cell IIe Microdissection Systems (Arcturus/Molecular Devices). These use a laser capture strategy that permits the visual inspection and editing of harvested cells for quality control before extraction of RNA (**Supplementary Methods**). For each cell type, we isolated at least 500–1,000 cells per biological replicate. This yielded 5–10 ng of total RNA, which we amplified into 20–50  $\mu\text{g}$  of mRNA-derived probe, as described below.

**RNA extraction and evaluation.** RNA was isolated using a Picopure kit (Arcturus/Molecular Devices) following the manufacturer's protocol. An aliquot of each RNA sample was measured fluorometrically (485 nm excitation, 535 nm emission) using RiboGreen reagent (Molecular Probes). RNA quality was checked in pre- and post-capture samples by evaluation on an Agilent 2100 Bioanalyzer using RNA-6000 Pico LabChips (Agilent Technologies) according to the manufacturer's instructions.

**RNA amplification and labeling.** To amplify cell-specific RNA into amounts sufficient for microarray probes, a T7 polymerase-based linear RNA amplification system (RiboAmpOA, Arcturus/Molecular Devices) was used for two rounds of amplification (**Supplementary Methods**). Because the amplification incorporates a primer that anneals to the poly(A) tail of mRNA, products have a bias that best represents the 3' end of template genes. This corresponds well to the 3' gene bias of the 70-mer oligonucleotides designed into the rice microarray platform to enhance gene hybridization specificity<sup>22</sup>. For microarray probes, the amplified RNA (aRNA) was labeled with 5-(3-amino allyl)-UTP (aa-UTP, Ambion) during the last step of *in vitro* transcription (in the second round of amplification).

**Preparation of labeled aRNA probes and microarray hybridization.** Purified aa-UTP-labeled aRNAs were resuspended in 0.1 M  $\text{NaHCO}_3$ . The aRNA probe

was further fluorescently labeled by conjugating the monofunctional Cy3 or Cy5 dye (GE Healthcare) to the amino allyl functional groups. Coupling was allowed to proceed at 22–24 °C for 45–60 min, and the labeling reaction was then stopped by ethanolamine. The fluorescent dye-labeled probe was separated from unincorporated dye and concentrated to a final volume of 7  $\mu\text{l}$  for hybridization using a Microcon YM-30 filter (Millipore). Hybridizations and scanning were performed as described in **Supplementary Methods**.

**Microarray experimental design and dye effect assessment.** Microarray experiments for 40 cell types were organized using a common reference statistical design to ensure that data for all cells could be normalized for comparison<sup>23</sup>. In this design, the hybridization for each cell type sample included a second probe made from an RNA sample that served as a common reference, identically aliquoted from a single preparation for all hybridizations and including dye swaps for the two probe labels. The common reference RNA for all hybridizations was prepared from rice suspension cultures, selected as an undifferentiated cell type that expresses a relatively high percentage of its genome<sup>22</sup>. Although this is an abundant source of RNA, the common reference RNA was amplified and labeled in identical fashion to other probes. At least four independent biological replicates for each cell type were collected and profiled; only microarray data sets with  $r \geq 0.9$  were retained. We included a dye swap in which each cell type sample and the common reference were labeled with Cy3 and Cy5, respectively, for two replicates, and labeled with Cy5 and Cy3, respectively, for the other two replicates. Quality parameters are described in **Supplementary Methods**.

**Data extraction and normalization.** The rice microarray platform consists of 60,000 spots on two separate slide arrays, including 16 repeats of 12 oligonucleotides on each array representing the negative controls<sup>22</sup>. We used the 97th percentile of the intensities from the negative controls as the threshold for significant expression. For global descriptions of transcriptome expression in the text, we included only genes above the threshold of expression in three out of four replicates. With the 97% cutoff, only five or fewer of the 165 negative controls were above the threshold. We found that a more stringent threshold eliminated biologically significant signals (see **Supplementary Methods** for further detail). For this cell type transcriptome atlas, normalization methods that assume a common distribution of expression for most genes across all experimental conditions were not suitable, as only a small proportion of genes were expressed across all 40 cell types, and because most genes are expected to have distinct expression patterns across cell types. To address these issues, we developed an iterative algorithm that alternates between expression ratio estimation and dye effect normalization to normalize the relative expression of expressed genes over the four dye-swap replicates in each experiment (see **Supplementary Methods** for further details on the procedure). Raw data, including array scan images, and normalized data were deposited with TIGR, and all normalized data sets are available from the project website (see URLs section of Methods).

**Validation of microarray data by RT-PCR.** We performed RT-PCR measurements for a set of genes selected on the basis of their relatively broad distribution of expression among all cell types (**Supplementary Table 2** and **Supplementary Methods**). Each gene was analyzed from at least three different samples for each cell type.

**Identification of chromosomal clusters of cell-specific genes.** Chromosomal clusters were identified using a simplified version of SaTScan<sup>24</sup>, a software program for identifying statistically significant clusters over time and space (**Supplementary Methods**).

**Cluster analysis.** Cluster analysis was applied to all genes showing cell-specific expression (as described above) in at least one cell type. Genes missing data in any experiment were removed. Relative expression, represented by  $z$  scores, were subjected to cluster analysis. For hierarchical clustering, a Pearson correlation was used to compute similarities, and the complete linkage clustering algorithm was used. Cluster analysis was performed and visualized using CLUSTER/JAVA TREEVIEW<sup>25,26</sup>. Several alternative clustering strategies produced the same hierarchical cellular relationships shown in **Figure 1**.

**GO enrichment analysis.** GO annotations from TIGR (version 5) and Gramene (12 January 2008 version) were combined. The enrichment within a query list was tested in comparison with all genes on the microarray. The web-based program Gostat<sup>27</sup> was used to find statistically overrepresented GO terms within groups of cell-enriched transcripts using the custom-defined GO annotation described above and in **Supplementary Methods**.

**Metabolic pathway analysis.** We classified biochemical pathway genes following the RiceCyc database (see URLs section below), which is based on MetaCyc pathway collections. Selected pathways (**Fig. 2b**) are shown as glyphs consisting of nodes, which represent metabolites, and lines, which represent reactions. Expression-level change of each reaction is shown in a color relative to the expression.

**Identification of promoter motifs enriched in genes encoding cell-specific transcripts.** The sequences 1 kb upstream of annotated translation start sites were retrieved from the TIGR database for cell-specific genes with a *z* score > 2 in a particular cell type. To find all enriched *cis*-elements, both DNA strands were searched using three independent methods detailed in **Supplementary Methods**. After aligning enriched elements, we manually identified common motifs found by all search approaches.

**Identification of housekeeping genes.** Expression values were used as a filter to identify putative housekeeping genes using a statistical strategy<sup>11</sup> with the following criteria: (i) ubiquitous expression and no more than two cell-type nulls, and (ii) uniform expression and a difference between the highest third and lowest third of log-transformed normalized values of < 2 (248 genes), < 2.5 (854 genes) or < 3.0 (1,588 genes). We identified 248 genes with a difference of 2, 854 genes with a difference of 2.5 and 1,588 genes with a difference of 3. Putative housekeeping genes were grouped according to the median of the log-transformed normalized expression values among samples: group I, 0–7.5; group II, 7.5–10; group III, 10–12.5; group IV, 12.5–15; group V, > 15.

**Hormone-related genes.** Annotation and functional information for hormone-related genes, including receptors, was obtained from the TIGR rice database (version 5) and from published sources<sup>28–30</sup>. Compiled lists of hormone-related genes are available at the URL listed below.

**RICEATLAS database and web interface.** The RICEATLAS website was implemented as detailed in **Supplementary Methods**. The website provides a variety of tools to view, search and analyze the data. In addition to providing details about the project, the site allows users to search through our data using a variety of criteria. Statistical tests can be performed on the data at the level of both the gene and the pathway. We have implemented a *t*-test and one-way analysis of variance (ANOVA) to find significant genes and a Fisher exact test and global test to locate pathways of interest.

**URLs.** Rice Atlas database and information, images of rice cell types, comparisons of ortholog expression patterns in *O. sativa* and *A. thaliana*, lists of housekeeping genes and compiled lists of hormone-related genes: <http://plantgenomics.biology.yale.edu/riceatlas/>; Rice Expression Database: <http://red.dna.affrc.go.jp/RED/>; gene annotations: <http://www.TIGR.org>; RiceCyc database: <http://www.gramene.org/pathway/>.

**Accession codes.** NCBI Gene Expression Omnibus (GEO): microarray data have been submitted under accession number GSE13161.

*Note: Supplementary information is available on the Nature Genetics website.*

#### ACKNOWLEDGMENTS

We thank N. Li for valuable database advice and assistance and P. Wu (Zhejiang University) for the rice root micrograph in **Figure 3**. This work was supported by US National Science Foundation Plant Genome Program grant DBI-0325821 to T.N., X.-W.D. and H.Z. T.L. and M.C. were supported in part by Peking-Yale Monsanto Fellowships.

#### AUTHOR CONTRIBUTIONS

T.N., X.-W.D. and H.Z. conceived and oversaw the research. S.L.T., N.G. and T.L. performed cell isolations, RNA isolations and informatic analysis. Y.J., H.Z. and

L.M. performed microarray hybridizations and informatic analysis. T.C., N.K.C. and M.C. performed cell and RNA isolations. N.S. designed and performed statistical methods for data processing and analysis. M.H. designed and implemented the atlas database and analytical tools. T.N. prepared the manuscript, with assistance from all coauthors.

Published online at <http://www.nature.com/naturegenetics/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Yu, H., Xia, Y., Trifonov, V. & Gerstein, M. Design principles of molecular networks revealed by global comparisons and composite motifs. *Genome Biol.* **7**, R55 (2006).
2. Birnbaum, K. *et al.* A gene expression map of the *Arabidopsis* root. *Science* **302**, 1956–1960 (2003).
3. Brady, S.M. *et al.* A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**, 801–806 (2007).
4. Nelson, T., Gandotra, N. & Tausta, S.L. Plant cell types: reporting and sampling with new technologies. *Curr. Opin. Plant Biol.* **11**, 567–573 (2008).
5. Nelson, T., Tausta, S.L., Gandotra, N. & Liu, T. Laser microdissection of plant tissue: what you see is what you get. *Annu. Rev. Plant Biol.* **57**, 181–201 (2006).
6. Ishiwatari, Y. *et al.* Rice phloem thioredoxin h has the capacity to mediate its own cell-to-cell transport through plasmodesmata. *Planta* **205**, 12–22 (1998).
7. Carland, F.M. & Nelson, T. Cotyledon vascular pattern2-mediated inositol (1,4,5) triphosphate signal transduction is essential for closed venation patterns of *Arabidopsis* foliar organs. *Plant Cell* **16**, 1263–1275 (2004).
8. Petricka, J.J. & Nelson, T.M. *Arabidopsis* nucleolin affects plant development and patterning. *Plant Physiol.* **144**, 173–186 (2007).
9. Zhan, S., Horrocks, J. & Lukens, L.N. Islands of co-expressed neighbouring genes in *Arabidopsis thaliana* suggest higher-order chromosome domains. *Plant J.* **45**, 347–357 (2006).
10. Hurst, L.D., Pal, C. & Lercher, M.J. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310 (2004).
11. Lee, S., Jo, M., Lee, J., Koh, S.S. & Kim, S. Identification of novel universal housekeeping genes by statistical analysis of microarray data. *J. Biochem. Mol. Biol.* **40**, 226–231 (2007).
12. Yang, G.X. *et al.* Microarray analysis of brassinosteroid- and gibberellin-regulated gene expression in rice seedlings. *Mol. Genet. Genomics* **271**, 468–478 (2004).
13. Yazaki, J. *et al.* Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and *Arabidopsis*. *Physiol. Genomics* **17**, 87–100 (2004).
14. Liu, X. *et al.* A G protein-coupled receptor is a plasma membrane receptor for the plant hormone abscisic acid. *Science* **315**, 1712–1716 (2007).
15. Razem, F.A., El-Kereamy, A., Abrams, S.R. & Hill, R.D. The RNA-binding protein FCA is an abscisic acid receptor. *Nature* **439**, 290–294 (2006).
16. Shen, Y.Y. *et al.* The Mg-chelatase H subunit is an abscisic acid receptor. *Nature* **443**, 823–826 (2006).
17. Abe, H. *et al.* *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **15**, 63–78 (2003).
18. Edwards, D., Murray, J.A. & Smith, A.G. Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in *Arabidopsis*. *Plant Physiol.* **117**, 1015–1022 (1998).
19. Lotan, T. *et al.* *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195–1205 (1998).
20. Coustry, F., Sinha, S., Maity, S.N. & Crombrugghe, B. The two activation domains of the CCAAT-binding factor CBF interact with the dTAF110 component of the *Drosophila* TFIID complex. *Biochem. J.* **331**, 291–297 (1998).
21. Gusmaroli, G., Tonelli, C. & Mantovani, R. Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. *Gene* **283**, 41–48 (2002).
22. Ma, L. *et al.* A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. *Genome Res.* **15**, 1274–1283 (2005).
23. Churchill, G.A. Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **32** Suppl, 490–495 (2002).
24. Kulldorff, M., Rand, K., Gherman, G., Williams, W. & DeFrancesco, D. SaTScan v2.1: software for the spatial and space-time scan statistics. *US National Cancer Institute* <<http://www.cancer.gov/prevention/BB/SaTScan.html#>> (1998).
25. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
26. Saldanha, A.J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
27. Beissbarth, T. & Speed, T.P. Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464–1465 (2004).
28. Du, L. *et al.* The two-component signal system in rice (*Oryza sativa* L.): a genome-wide study of cytokinin signal perception and transduction. *Genomics* **89**, 697–707 (2007).
29. Ito, Y. & Kurata, N. Identification and characterization of cytokinin-signalling gene families in rice. *Gene* **382**, 57–65 (2006).
30. Jain, M. *et al.* Structure and expression analysis of early auxin-responsive Aux/IAA gene family in rice (*Oryza sativa*). *Funct. Integr. Genomics* **6**, 47–59 (2006).