

Root genomics: towards digital *in situ* hybridization

Ben Scheres, Henk van den Toorn and Renze Heidstra

Address: Department of Molecular Cell Biology, Utrecht University, Padualaan 3, 3584 CH Utrecht, The Netherlands.

Correspondence: Ben Scheres. E-mail: b.scheres@bio.uu.nl

Published: 27 May 2004

Genome Biology 2004, **5**:227

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/6/227>

© 2004 BioMed Central Ltd

Abstract

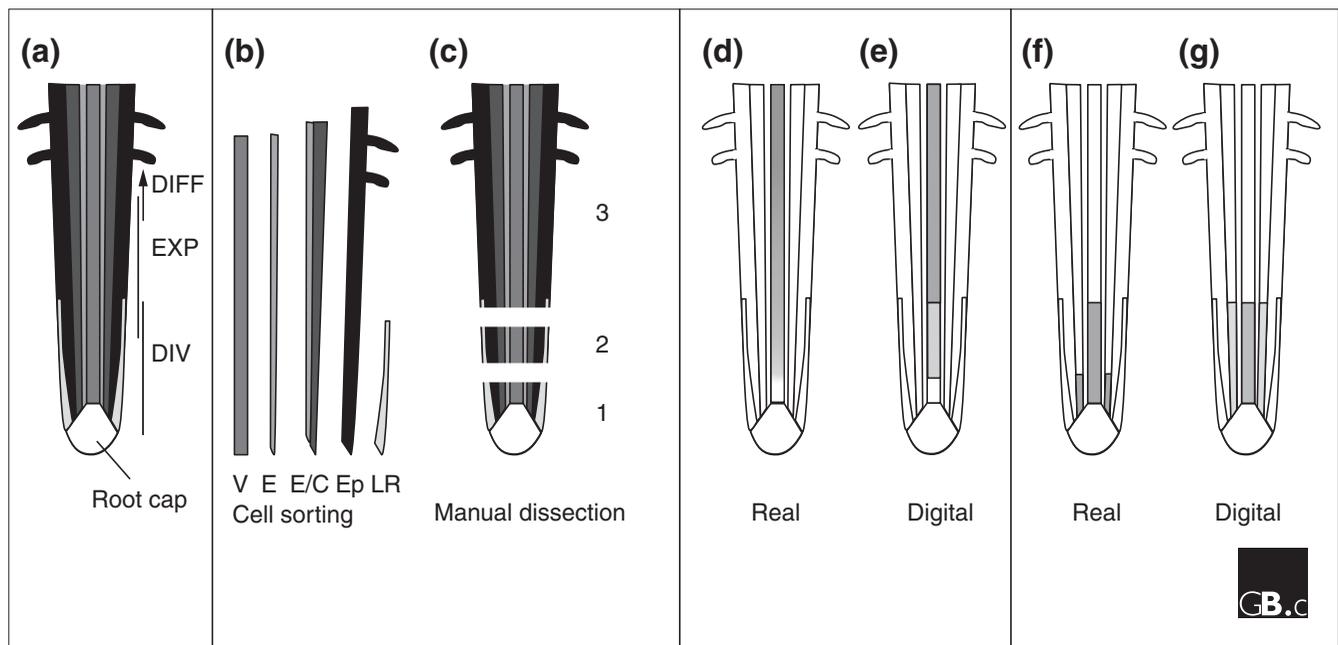
Separation of cell types and developmental stages in the *Arabidopsis* root and subsequent expression profiling have yielded a valuable dataset that can be used to select candidate genes for detailed study and to start probing the complexities of gene regulation in plant development.

Tracking developmental changes in gene expression

The availability of genome-wide expression analysis tools allows one to investigate the details of transcriptional regulation during development. Clustering methods can be used to group genes whose expression varies in a similar way in response to developmental changes. Such clustering methods can reveal two major trends. First, they can reveal groups of genes that are co-regulated, and therefore suggest which genes function together during a given developmental process. Second, clustering methods can reveal which conditions resemble each other, pointing out similarities - or dissimilarities - in developmental states that might not be obvious otherwise. Two major developmental parameters for analysis by gene-expression profiling are progression in time ('developmental stage') and tissue, region or cell-type specificity. Previous studies of gene expression during the development of multicellular organisms have mostly emphasized either the developmental stage or the cell-type aspect. For example, clusters of genes co-expressed during the entire life cycle have been defined in *Caenorhabditis elegans* [1], and changes at the transition from cell proliferation to cell differentiation have been described for the *Drosophila* eye [2]. Another *C. elegans* study emphasized cell-type-specific gene-expression programs [3]. Both temporal and spatial aspects of gene expression have been analyzed by transcript profiling of the slime mold *Dictyostelium discoideum*, an organism in which cell aggregation leads to a multicellular structure with two different mature cell types [4]. Recently,

Birnbaum *et al.* [5] have conducted a global gene-expression analysis of a more complex mix of cell types at three developmental stages in the small weed *Arabidopsis*, and have generated a digital reconstruction of gene expression in the root - a 'digital *in situ* hybridization'.

Higher plants, like animals, develop from a single cell, but the majority of the plant body derives from the post-embryonic activity of clusters of stem cells and their mitotically active daughters, the meristems. After dividing, meristematic cells displace daughter cells that subsequently differentiate at a distance from the mitotic cell pool. This is a particularly regular process in the *Arabidopsis* root (Figure 1a) [6], and because of this regularity cells of different developmental stages occupy defined regions of cell division, cell expansion and cell differentiation. In the radial dimension, the root meristem extends concentrically arranged tissues that represent the root-specific versions of the main plant tissues: epidermal, ground (endodermal and cortical) and vascular tissue. Over the years, a number of genes have been identified that are important for pattern formation, cell cycle and cell growth, and hormone signaling; and these genes are beginning to provide an understanding of the developmental processes that occur in the root meristem [7]. But much more information is needed if we are to identify the details of the regulatory network(s) that determines cell identity, directional cell division, polar expansion and growth parameters. Obviously, detailed knowledge of the transcript localization for (nearly) all genes in an organ is an important step towards achieving this goal.

**Figure 1**

Dissection of gene-expression domains in the *Arabidopsis* root. **(a)** Schematic overview of the root. DIV, cell division zone; EXP, zone of rapid cell expansion; DIFF, zone of cell differentiation. **(b)** Tissue and cell types as sorted by fluorescence-activated cell sorting (FACS) in the study by Birnbaum *et al.* [5]. V, (pro-)vascular cells; E, endodermis; E/C, endodermis and cortex; Ep, epidermis; LR, lateral root cap. **(c)** Manually dissected regions, also used in [5]. **(d)** Gene-expression patterns that are distributed in a graded manner through the developmental stages become discrete in **(e)** the 'digital *in situ*' representation. **(f)** The expression pattern of genes expressed in distinct zones that differ per tissue type becomes averaged in **(g)** the digital version throughout the tissues and stages.

Separation of cell types and developmental stages

Several approaches have been designed for obtaining RNA from specific stages or cell types. Stage-specific promoters can be fused to the green fluorescent protein (GFP), and cell populations can be purified by fluorescence-activated cell sorting (FACS) of trypsin-dissociated cells [2]. Alternatively, cell-type-specific expression of epitope-tagged RNA-binding proteins can be used to enrich mRNA [3]. Laser-assisted microdissection of specific cells is also possible [8,9]. RNA from specific developmental stages or tissue regions obtained in these ways can be analyzed by microarray technology or serial analysis of gene expression (SAGE). The recent study from the Benfey group [5] used oligonucleotide chips to analyze gene expression in *Arabidopsis* roots; they first dissected out the major tissues by enzymatically dissociating cells (protoplasting) and doing FACS analysis of transgenic lines expressing GFP under region- or cell-type-specific promoters (Figure 1b). It may perhaps seem tricky to enzymatically digest cell walls and then sort protoplasts, asking them to maintain cell-fate- or region-specific expression patterns for 1.5 hours. After all, plant biologists are used to the flexibility of cell-fate determination in the plant kingdom, with the - somewhat overstated - textbook dogma that plant cells are totipotent and maintain their identity only in the context of the organism. Yet, amazingly, this approach proved successful. Only a minor set of

genes appeared to be induced by protoplasting and sorting, and these were removed from the analysis.

Hence, Birnbaum *et al.* [5] were able to isolate RNA from GFP-expressing, sorted vascular, ground-tissue and epidermal cells (see Figure 1b,c) and hybridize it to the Affymetrix ATH1 GeneChip, which has probes for approximately 22,000 *Arabidopsis* genes, covering about 90% of the genome. In a separate experiment, manual dissection of three developmental zones allowed the authors to determine the relative level of expression of each gene in zones roughly representing three different stages: cell proliferation, cell expansion and cell differentiation (Figure 1c). For every gene, this percentage was then superimposed on the expression values per tissue or cell type. Validation experiments using both previously documented and new genes confirmed that this method gives reliable expression data for the majority of genes.

While the starting dataset is already impressive, the method used lends itself to future improvements that will further enhance the resolution. First, by means of bootstrapping, the promoters of candidate cell- or region-specific genes that emerge from the first analysis can now be used to refine the set of GFP lines that are used for cell sorting. In the future it is likely to be possible to sort all the different root cell types separately. Second, the stage-specific and the tissue-specific

gene-profiling data are currently combined by calculation, which works best if genes have sharp expression transitions and the same distribution over the three developmental stages in every cell type, which will not be the case for all genes. For example, a gene with a graded transcript distribution, or a gene whose stage-dependent transcription differs from one tissue to another, will not be recognized as such in the current dataset (Figure 1d-g). In the future this limitation can be overcome by sorting cell types from separately dissected stages. Another option is to combine stage- and cell-type-specific markers, and to sort cells that possess both.

Using expression maps to generate hypotheses

The current dataset of gene expression in the root [5] provides a rich resource for those interested in plant development. Cell-type-specific expression of each researcher's favorite gene in the root suggests a starting point for searching for mutant phenotypes of interest, and the ease with which cellular details of phenotypes can be visualized in the root can facilitate detailed analysis of genes that may first be identified from studies in other organs. For those interested in root development itself, functional redundancy can now be overcome more easily by selecting homologs of genes that have overlapping expression profiles. Potential targets for known transcription factors can be pre-selected or validated because they should be co-expressed in at least a subset of the cell types that express the transcription factor of interest. The mRNA enrichment obtained by sorting can be exploited to enhance the sensitivity of detecting transcriptional differences in mutants, after gene induction experiments or after drug treatments. Map-based cloning of genes can be accelerated because expression patterns matching with region-specific root phenotypes can be selected when mapping intervals are still large. The excellent *Arabidopsis* resources for the recovery of insertion mutants [10], and mutants induced by ethylmethane sulfonate (EMS) through the TILLING procedure [11] provide useful and rapid follow-up resources for such a candidate-gene approach. In all these, and probably more, applications, the dataset is used as a starting point for further analysis.

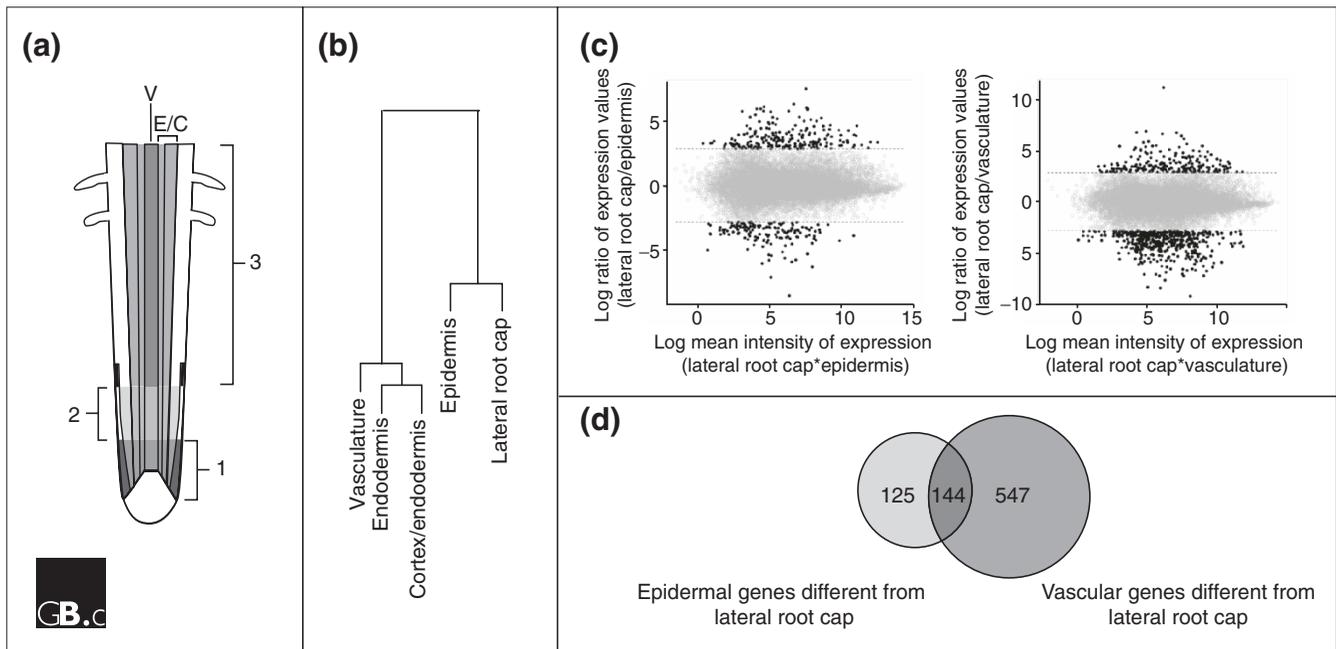
A major question that remains to be answered is the extent to which complex gene-expression maps reveal underlying regulatory features. Many computational tools can be used to cluster gene-expression data into meaningful groups, and the tool chosen largely determines what information is highlighted from the dataset [12]. In the *Drosophila* eye, hierarchical clustering using expression data and gene function as input revealed a cluster with cell-cycle and cell-growth regulators enriched in proliferating cells, a signaling and adhesion cluster in early-stage differentiating cells, and a cluster enriched in transcription factors in the mixed population of photoreceptor and cone cells [2]. In the slime mold, aggregation of single-celled amoebae leads to a dramatic morphological change, giving rise to a multicellular organism with two mature cell types. In this case, a striking amount of gene

regulation could be observed by fitting all differentially expressed genes to a hypothetical gene-induction curve; and the similarities between expression profiles for all genes in each developmental stage revealed that the transition from unicellular to multicellular stages was accompanied by a dramatic change in gene-expression programs involving changes in around 25% of all transcripts. Purification of cell types and their precursors, subsequent microarray analysis and fitting the data to functions that represent particular kinds of cell-type enrichment, revealed the existence of clear cell-type-specific clusters [4].

Birnbaum *et al.* [5] used binary coding, principal component analysis and *k*-means clustering to find dominant expression patterns among the 5,712 differentially expressed genes (defined as having more than a four-fold difference between any two conditions) in roots (Figure 2a). These clusters show up on a visual representation of all expression data. The largest cluster comprised around 30% of these genes and showed upregulation in the proliferation stage in all cell types. This cluster contained a majority of genes involved in the cell cycle and nuclear organization - reminiscent of the proliferation-associated gene cluster in fly eyes. Also apparent from the clustering was that a large class of genes (approximately 10%) is specifically upregulated in differentiated vascular tissue, consistent with the presence of several very different cell types within this tissue. When the gene content was analyzed, several functional categories - those involved in hormonal signaling pathways, for example - appeared over-represented in some clusters compared to others [5]. Although this statistical over-representation might indicate a higher importance of certain hormone pathways in specific regions, it is as yet unclear whether statistical significance implies biological significance.

The major clusters found by Birnbaum *et al.* [5] reveal some other trends in root development that raise interesting questions. For example, consistent with the presence of mature layers of lateral root cap surrounding the meristem at close proximity to the tip, it is not surprising that genes enriched in the lateral root cap appear in the proliferation stage. Interestingly however, vascular and ground-tissue cells appear to achieve their tissue-specific expression patterns at a larger distance from the apex than the epidermal cells do. It is not clear why genes enriched in epidermal cells would be switched on at closer proximity to the stem cells than genes enriched in vascular cells, while overt differentiation characteristics in both tissues appear at roughly similar distances from the apex. A simple explanation may be that early cell-type-specific genes in the vasculature may be diluted beyond detection because, in contrast to the epidermis and endodermis, the vascular tissue is a mixture of cell types.

A rich resource like the root expression map opens up numerous possibilities for data analysis. For example, 'similarity' calculations like those used in *Dictyostelium* [4]

**Figure 2**

Global analysis of gene expression in the root. **(a)** Major clusters of co-expressed genes called localized expression domains (LEDs) from the analysis by Birnbaum *et al.* [5]. V, vascular tissue; E/C, endodermis and cortex. 1, 2 and 3 refer to the dissection zones in Figure 1c. **(b-d)** Our own analysis of the data from [5]. **(b)** A similarity tree calculated from the data in [5] using Euclidian distance with complete linkage. For all five tissues, all genes were taken as coordinates, resulting in five points in a multidimensional space. The Euclidian geometric distance between these points was calculated. To obtain the clustering, the points closest in space (vasculature and cortex/endodermis) were defined as the first cluster. All other points are subsequently added to this cluster based on the point furthest away inside the cluster. **(c)** Two cluster diagrams showing the similarity between tissue types using the Canberra similarity measure with complete linkage (see [16]). For all five tissues, all genes were compared using a similarity measure between experiments. Cell types were compared using log ratio of expression values ($m = \log_2(\text{tissue a}/\text{tissue b})$) versus log mean intensity of expression ($a = \log_2(\text{tissue a} * \text{tissue b})/2$) plots using the R statistical language [17,18]. After transforming the data, linearity was corrected using the Loess function, and further analysis was done on residuals. The two tissues resembling each other most (lateral root cap and epidermis) and least (vasculature and lateral root cap) in the dendrograms are analyzed. The threshold for differential expression is three times the standard deviation of the experiment with the least variance (lateral root cap versus epidermis) on both scatter plots (dotted lines). Differentially regulated genes are shown as filled circles outside the dotted lines. **(d)** Numbers of genes differentially regulated under these restrictions shown as a Venn diagram.

reveal expression profiles of vascular and ground-tissue cells to be much more similar to each other than to the outer epidermal and lateral root cap cells (Figure 2b,c). Selection and sorting for cell-type-specific expression, on the other hand, provides an estimate for the critical differences between cell types (Figure 2d). By viewing the data in these and other ways, different aspects of the dataset are highlighted, each providing useful new insights.

With the first version of the root digital *in situ* hybridization map at hand, more regularities within the datasets can be explored. Candidate tissue- or stage-specific transcription factors can be analyzed for direct or indirect roles in the expression of their co-regulated genes, which might explain at least part of the data as resulting from the activity of a transcription-factor network. How easy this is will depend on how many layers of regulation at the post-transcriptional level are responsible for the ultimate distribution of mRNAs in the root, and how many of the transcriptional differences are pre-established by factors no longer expressed at the post-embryonic stage.

It is to be expected that, as new tissue- or stage-specific datasets are provided from other regions of *Arabidopsis* (see, for example, [13-15]), the root data can be inspected using many additional filters. For example, truly root-specific genes can be separated from those that are expressed in other organs, creating interesting new groups such as root proliferation-stage genes that are also expressed in the shoot apical meristem. While much work remains to be done to refine the root expression map and to integrate it with other expression data, the initial work presented by Birnbaum *et al.* [5] opens the doors to these possibilities and others yet to be foreseen.

References

- Hill AA, Hunter CP, Tsung BT, Tucker-Kellog G, Brown EL: **Genomic analysis of gene expression in *C. elegans***. *Science* 2000, **290**:809-812.
- Jasper H, Benes V, Atzberger A, Sauer S, Ansorge W, Bohmann D: **A genomic switch at the transition from cell proliferation to terminal differentiation in the *Drosophila* eye**. *Dev Cell* 2002, **3**:511-521.

3. Roy PJ, Stuart JM, Lund J, Kim SK: **Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans***. *Nature* 2002, **418**:975-979.
4. Van Driessche N, Shaw C, Katoh M, Morio T, Sugang R, Ibarra M, Kuwayama H, Saito T, Urushihara H, Maeda M, et al.: **A transcriptional profile of multicellular development in *Dictyostelium discoideum***. *Development* 2002, **129**:1543-1552.
5. 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.
6. Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts R, Scheres B: **Cellular organisation of the *Arabidopsis thaliana* root**. *Development* 1993, **119**:71-84.
7. Scheres B, Benfey P, Dolan L: **Root development**. In *The Arabidopsis Book*. Edited by Somerville CR and Meyerowitz EM. Rockville, MD: American Society of Plant Biologists; 2002. [<http://www.aspb.org/downloads/arabidopsis/scheres.pdf>]
8. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA: **Laser capture microdissection**. *Science* 1996, **274**:998-1001.
9. Asano T, Masumura T, Kusano H, Kikuchi S, Kurita A, Shimada H, Kadowaki K: **Construction of a specialized cDNA library from plant cells isolated by laser capture microdissection: toward comprehensive analysis of the genes expressed in the rice phloem**. *Plant J* 2002, **32**:401-408.
10. **World-wide Arabidopsis Reverse Genetic Stocks** [http://www.arabidopsis.org/info/2010_projects/Reverse_Genetic_Stocks.jsp]
11. **Arabidopsis TILLING project** [<http://tilling.fhcrc.org:9366/>]
12. Quackenbush J: **Computational analysis of microarray data**. *Nat Rev Genet* 2001, **2**:418-427.
13. Brandt S, Kloska S, Altmann T, Kehr J: **Using array hybridization to monitor gene expression at the single cell level**. *J Exp Bot* 2002, **53**:2315-2323.
14. Ruuska SA, Girke T, Benning C, Ohlrogge JB: **Contrapuntal networks of gene expression during *Arabidopsis* seed filling**. *Plant Cell* 2002, **14**:1191-1206.
15. Che P, Gingerich DJ, Lall S, Howell SH: **Global and hormone-induced gene expression changes during shoot development in *Arabidopsis***. *Plant Cell* 2002, **14**:2771-2785.
16. Legendre L, Legendre P: *Numerical Ecology*. Amsterdam: Elsevier; 1983.
17. R Development Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna; 2003.
18. **The R project for statistical computing** [<http://www.R-project.org>]