The transcriptional complexity of barley development

A. Druka, I. Druka, D.F. Marshall, J. McNicol & R. Waugh*

Since the late 1990s, the establishment of microarray technology has allowed the abundance of messenger RNAs (mRNAs: the intermediates between genes and proteins) derived from thousands of genes to be measured in a single experiment. The ability to monitor how the majority of the mRNA transcripts in a cell or tissue change in abundance over time, in different tissues or in response to environmental cues is immensely powerful and helps identify networks of genes that are co-regulated or that are specific to individual tissues or cell types, even if the functions of the genes are not currently known. Microarray data is thus informationally very rich, and using relatively simple designs such experiments have yielded the first insights into the gene expression networks that control a range of biological processes.

Over the last three years we have collaborated with the world barley genomics community and a commercial vendor to develop the technically robust Barley1 microarray that exploits GeneChip® technology. It can be used to assess the abundance of mRNA transcripts from over 20,000 different barley genes at one time. We have used this system to examine mRNA abundance in a range of different barley tissues representing the major stages in the development of the barley plant. The tissues are shown in Fig. 1A. This developmental experiment has provided insight into how mRNA transcript population complexity changes in different tissues during development and will act as a reference for future studies in barley and other grass species including wheat, rice and maize.

We investigated mRNA transcript population complexity by measuring for each tissue the number of genes that were expressed, the level at which they were expressed and how that level changed between the tissues examined. Looking at the complete mRNA population as a whole (termed the 'transcriptome') we found that highly differentiated tissues such as anthers

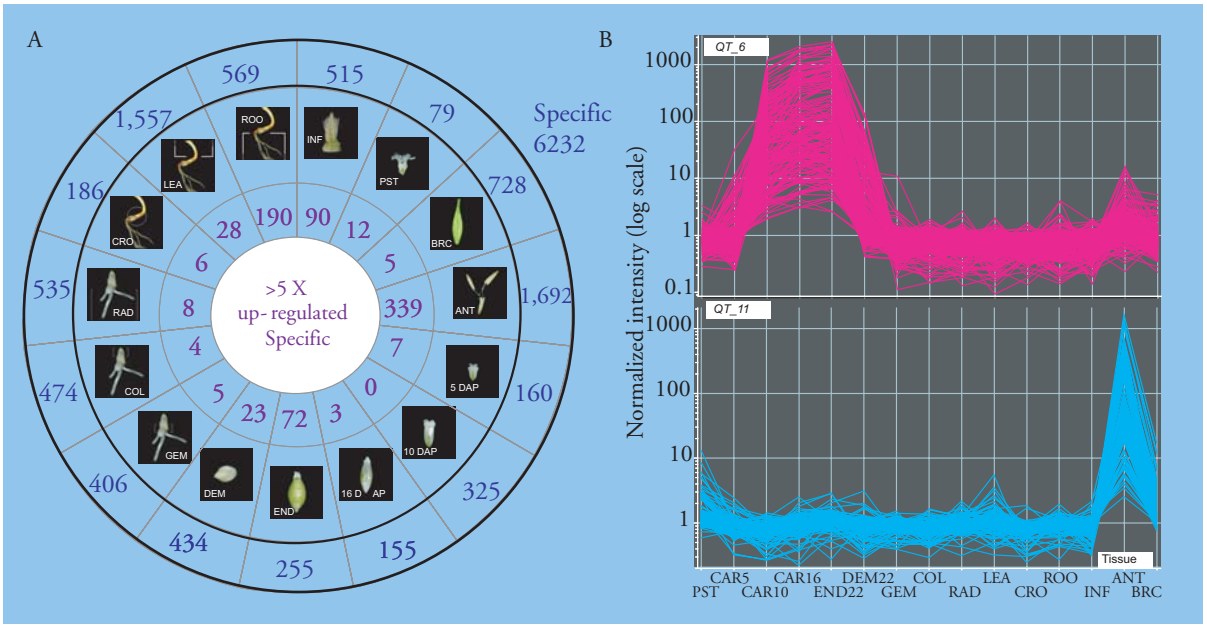

Figure 1 (A) Tissues we used to investigate transcriptional complexity of barley development. Respective Cereal Plant Anatomy and Plant Growth stage terms were according to ontologies developed by Gramene. The following are the tissue abbreviations: INF, inflorescence; PST, pistil; BRC, bracts (lemma, palea and glumes); ANT, anthers; CAR5, caryopsis 5; CAR10, CAR16; END22, caryopsis without embryo; DEM22, embryo; GEM, mesocotyl; COL, coleoptile; RAD, seminal root; CRO, crown; LEA, leaf (partially shown); ROO, root (partially shown).
(B) Two examples of clusters of genes that show similar patterns of gene expression across all 15 tissues examined. The top panel shows genes that are predominantly expressed during stages of seed development. The bottom panel identifies genes that are specifically expressed only in the developing anther tissues. In both cases a large number of 'unknown' genes cluster together with those of established function.

*This report is the summary of an international collaborative project:.
Druka,A., Muehlbauer,G., Druka,I., Caldo,R., Baumann,U., Rostoks,N, Schreiber,A, Wise,R., Close,T, Kleinhofs,A, Graner,A, Schulman,A., Langridge,P, Sato,K, Hayes,P, McNicol,J, Marshall,D & Waugh,R. (2004) submitted

expressed fewer genes than tissues containing many or actively proliferating cell types with a high positive correlation (r=0.94) between the number of genes expressed in a tissue and the total amount of mRNA (as a fraction of the total RNA which includes the ribosomal RNA). These observations allowed us to conclude that the amount of mRNA in a tissue is largely determined by the number of unique transcripts in the population, rather than varying the abundance of an existing set.

By comparing the lists of genes that were either on or off in the different tissues, the experiment showed that a large number of genes are turned on or switched off during barley development. To look at underlying trends in the dataset we then applied two complementary statistical analyses. The first allowed us to identify genes that are specifically switched-on in individual tissues and the second to group genes that showed common patterns of gene expression across all fifteen tissues. Fig. 1A gives the result of the first analysis. The numbers associated with each of the tissues represent the number of genes that we detected that were specifically expressed, and the subset of these that were >5-fold more abundant, in the indicated tissue compared to any other. These numbers therefore reflect the uniqueness of the individual transcriptomes. Barley anthers for example possess the greatest number of 'specific' transcripts compared to any of the other tissues and the caryopsis at 16 days post anthesis the least. The genes encoding these specific transcripts represent a valuable collection of candidates that can be used to investigate in great detail the underlying biology of the sampled tissues.

Fig. 1B gives two examples of the second type of analysis and shows how genes can be 'clustered' together based on common patterns of abundance across the tissues. Examining the information associated with the genes included in these groups revealed that where specific functions were known, the different members of the co-regulated clusters were frequently functionally related. However it also showed that a considerable portion of the co-regulated clusters were made up of unknown or hypothetical proteins. As a result of this analysis we can putatively assign functional annotations to each of these genes based solely on mRNA transcript abundance-based functional correlations.

Further detailed analysis of this developmental mRNA transcript abundance dataset confirms that it contains information that reflects known biological patterns and that it has considerable descriptive and predictive biological potential. Free public access will soon be available and we predict that this dataset will act as both a reference and a catalyst for hypothesis driven research in barley and other grasses by providing robust and representative transcript abundance profiles of approximately 22,000 genes.