The genetics of gene expression

Arnis Druka, Elena Potokina*, Mike Kearsey*, Zewei Luo* and Robbie Waugh

***University of Birmingham**

Most important plant characteristics, such as grain quality or yield, are controlled by many genes and influenced by the environment. In plant breeding and genetics these are usually monitored in defined populations by recording a score for that characteristic on a sliding scale. Such measurements are both subjective and often distantly removed from the biological processes giving rise to them (e.g. yield). Environmental and other factors may also impart significant additional variation. The resulting imprecision restricts our ability to identify the location of the genes controlling the chatacteristic and the overall approach provides no information at all about the complex network of genes or gene-products which are responsible for determining why the measured phenotype is at one extreme or the other. Advances in molecular biology have made it possible to measure the abundance of thousands of gene products, known as transcripts, in a single experiment by using an approach termed 'microarray technology'. For example, in barley we use a microarray that measures the abundance of 23,000 different transcripts in a single experiment. The overall profile of variation in transcript abundance obtained from such an experiment represents a snap-shot of the important biological processes that are operating in the material being studied at the time of sampling. By correlation with classical trait scores it has been shown that transcript abundance can serve as a 'surrogate' (i.e. an indirect measurement) for classically recorded characters. For example, variation in transcript abundance has been used to identify 'susceptibility loci' for complex diseases like obesity, asthma and diabetes in mammals, and the overall approach is now being used to differentiate between healthy and different forms of cancerous cells.

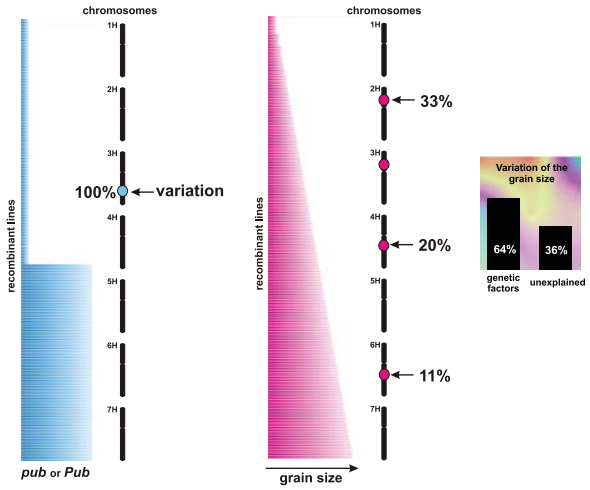

Figure 1 Phenotypic variation of Mendelian or binary traits can easily be explained by a single gene or locus. This contrasts with quantitative traits, where genetics behind phenotypic variation often is quite complex and usually not all of it can be attributed to the genetic factors. *Pub* and *pub* shown here as an example, are alleles of the unknown gene, that determine presence or absence of the hairs on the leaves. Such Mendelian or binary trait can easily be converted to the quantitative trait if number of hairs per individual line is counted.

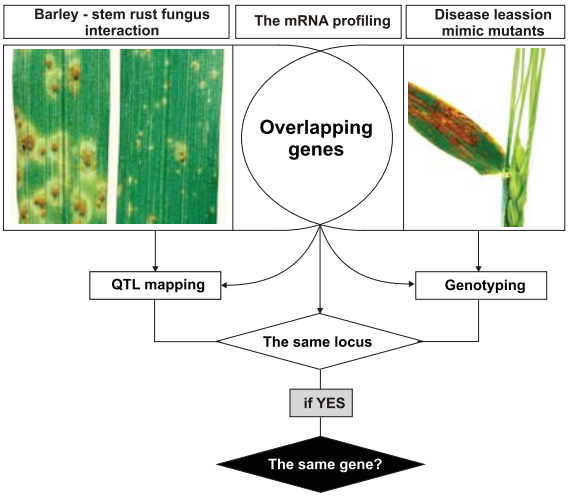

Figure 2 The strategy we use to identify candidate genes for complex traits. Three different components are employed; experimental population where the trait segregates, induced mutations–containing lines, that either have obvious phenotype related to that of segregating population (forward genetics approach), or mutations in the candidate genes (reverse genetics, TILLING). As a third component, mRNA abundance phenotypes of thousands of genes are used to link the first two.
In this example, we specifically address group of PCD-related genes that have eQTLs associated with those of partial resistance to the wheat stem rust fungus in barley. We also mapped the phenotype of one of the disease lesion mimic mutants to the same locus.

We quantified the relative abundance of 23,000 different transcripts in two different barley tissues from each of the individuals in a segregating population that has also been analysed for a wide range of classical phenotypic traits. We wanted to determine whether, like the mamalian studies, we could identify the major genetic determinants underlying classical characteristics while at the same time gain biological understanding of the networks of gene-products that underly their development. In our experiment, 12,738 of the barley genes on the chip recorded variation in transcript abundance that segregated in the population. By using standard genetic mapping approaches we were able to show that over a third of these 'transcript abundance' phenotypes behaved as single genes and as a result we were able to construct a high density map of the barley genome with each of the genes ordered along each of the chromosomes. We then performed correlation analysis with classically scored phenotype from the well studied interaction between barley and the wheat stem rust fungus *Puccinia graminis f. sp. t ritici* as a model. We showed that the approach successfully predicted the *Rpg1* resistance gene as a major contributor to resistance and highlighted several other candidate genes and loci that contribute to this interaction. We are now performing a detailed experimental analysis of one of the loci we identified and named Required for Puccinia resistance 2 (*Rpr2*) that putatively encodes a 'master regulator' of a process called programmed cell death. To allow both ourselves and remote users to investigate further correlations within the dataset to additional phenotypic traits we have deposited all the information in an online data analysis environment called the GeneNetwork (http://www.genenetwork.org/)