Most characteristics in living organisms show continuous variation, which suggests that they are controlled by multiple genes. Quantitative trait loci (QTL) analysis can identify the genes underlying continuous traits by establishing associations between genetic markers and observed phenotypic variation in a segregating population. The new high-throughput sequencing (HTS) technologies greatly facilitate QTL analysis by providing genetic markers at genome-wide resolution in any species without previous knowledge of its genome. In addition HTS serves to quantify molecular phenotypes, which aids to identify the loci responsible for QTLs and to understand the mechanisms underlying diversity. The constant improvements in price, experimental protocols, computational pipelines, and statistical frameworks are making feasible the use of HTS for any research group interested in quantitative genetics. In this review I discuss the application of HTS for molecular marker discovery, population genotyping, and expression profiling in QTL analysis.

**Keywords:** QTL analysis, plant genetics, next generation sequencing, genomics, eQTL analysis, RNA-seq

**INTRODUCTION**

For almost one century scientists have dissected the genetic architecture of quantitative traits in plants using Quantitative trait loci (QTL) analysis (Fisher, 1918). These analyses establish associations between genetic markers and the phenotypic variation of a quantitative trait in a segregating population. The techniques used to obtain markers and physiological phenotypes have been constantly improved through history (Schloter, 2004; Montes et al., 2007). Recently, the price drop of high-throughput technologies have allowed plant researchers to quantify the general abundance of transcripts, proteins, or metabolites in segregating populations (Kirst et al., 2005; Vuylsteke et al., 2005, 2006; Decook et al., 2006; Keurentjes et al., 2007; West et al., 2007; Lisec et al., 2008; Potokina et al., 2008; Drost et al., 2010). These studies show that there are multiple benefits in using “omic” technologies for QTL analyses, even when the goal is to characterize physiological phenotypic diversity. First, molecular phenotypes are the initial step toward the production of physiological phenotypes and its regulation underlies much of phenotypic diversity (Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008). Second, the availability of genome-wide information significantly increases the ability to identify candidate genes for QTLs (Jimenez-Gomez et al., 2010). Third, molecular traits measured at system scale allow estimation of the effect of QTLs in the genetic pathways of interest, or identification of additional gene networks altered by the loci responsible for the variation (Kliebenstein et al., 2006). Finally, molecular traits offer researchers a better understanding of how mutation drives physiological variation and what are the evolutionary forces acting at primary levels.

High-throughput sequencing, or HTS, allows the rapid and cost-effective generation of massive amounts of short sequences or reads (Metzker, 2010). The potential of this technology for mapping loci responsible for phenotypic differences in plants has already been demonstrated by identifying genes containing EMS-induced mutations in samples of pooled F2 individuals (Schneebberger et al., 2009; Austin et al., 2011). HTS technologies have been in the market for a few years, and new methods are being developed that will be cheaper, require less sample processing, and will produce more and longer reads (Munroe and Harris, 2010; Glenn, 2011; Niedringhaus et al., 2011). It is therefore clear that very soon HTS will be the tool of choice for QTL analyses. One important limiting factor remains to be eliminated: Data analysis. It requires long and computationally intensive pipelines that need to be customized for each particular experimental set up. An increasing number of new algorithms are constantly released to the community, and the debate on which pipelines return the most accurate results is still ongoing. Comparing, combining, and customizing these pipelines requires simple Unix or Linux commands and greatly benefits from knowledge in powerful statistical software such as R, and in scripting languages, such as Perl or Python (R Development Core Team, 2009). For non-bioinformaticians, integrated solutions with convenient interfaces are becoming popular both from collaborative open projects and companies (Blankenberg et al., 2010; Goëcks et al., 2010). A popular website that keeps an actualized list of the available software tools is www.seqanswers.com, where users and developers also discuss new technological advances and pipelines. In terms of the computational equipment required for HTS data analysis, the majority of tools are developed for Linux or Unix based systems. Although parts of the analysis can be performed in any modern computer, machines with dozens of gigabytes of RAM are recommended in cases where reference sequences form the species considered are available, or with hundreds if no reference exists. An alternative option that is likely to become popular is to rent storage and computing power in specialized centers, or “the cloud” (Stein, 2010).

Due to the fast improvement of HTS, this review intends only to capture a snapshot in time of the possibilities that it offers for
molecular marker discovery, genotyping, and molecular phenotyping in segregating populations of plants. This review has the purpose of helping researchers who have not incorporated this technology to their work to think about the requirements and possibilities of HTS. By no means this review refers to all available experimental designs or analysis tools, and the solutions proposed here are mere suggestions that will certainly soon be substituted by new and better ones. A guide map of the methods proposed in this review is depicted in Figure 1.

**LIBRARY PREPARATION**

Sample preparation protocols are continuously improved to use fewer amounts of biological material, be completed faster, and reduce the bias in their output. As an example, most current protocols allow multiplexing samples by adding a short sequence tag to all reads in a library, a convenient feature given the increasing numbers of reads produced per HTS run. The same companies that developed the HTS sequencers commercialize library preparation protocols optimized for the most common experimental designs. There are also kits from other companies that give comparable results and may be more cost efficient. Finally, many researchers are developing custom protocols to obtain specific information such as the transcribed strand in RNA-seq experiments, the rate of RNA degradation, or the positions occupied by RNA polymerases, just to name a few (Addo-Quaye et al., 2008; Core et al., 2008; German et al., 2008; Parkhomchuk et al., 2009).

**QUALITY CONTROL AND PRE-PROCESSING**

Assessing the quality of HTS reads includes detection of biases on base composition, base quality, and sample complexity. The quality of the sequences has an impact on the reliability of the biological interpretations resulting from the analysis (Dohm et al., 2008). Part of these biases are introduced by the sample preparation protocols (Schwartz et al., 2011), particularly during cDNA synthesis in RNA-seq experiments (Hansen et al., 2010; Li et al., 2010b) and PCR amplification (Aird et al., 2011). Additional biases are particular to each HTS technology (Smith et al., 2008; Quince et al., 2011) or specific to each run of the sequencers (Auer and Doerge, 2010).

After quality control it is usually necessary to pre-process the reads by trimming low quality nucleotides and adapter sequences. At this stage, foreign sequences such as vectors or DNA from organisms contaminating the samples can also be removed. Depending of the type of libraries sequenced further pre-processing may be needed, such as trimming poly A or poly T tails and terminal transferase tails in RNA-seq libraries. In cases where several libraries have been multiplexed, reads should be separated by their barcode.

Both quality control and pre-processing can be easily performed with basic scripts written in Perl (Bioperl), R (Bioconductor), or Python (Biopython; Stajich et al., 2002; Gentleman et al., 2004; Cock et al., 2009; R Development Core Team, 2009). For non-programmers, there are some convenient tools that can carry out all or some of these tasks (FastQC, 2008; FASTX-Toolkit, 2009; Blankenberg et al., 2010; Falgueras et al., 2010; Goecks et al., 2010; Cutadapt, 2010; Schmieder et al., 2010; Schmieder and Edwards, 2011).

**MOLECULAR MARKER DISCOVERY**

Depending on the availability of a reference sequence short reads will be aligned or de novo assembled using one of the multiple tools available. There are a number of recent articles that compare the most popular algorithms and software available for these purposes (Bao et al., 2011; Lin et al., 2011; Ruffalo et al., 2011). Please note that the methods proposed below are directed to developing molecular markers for QTL analysis and not to...
identify the mutation underlying the QTL, which requires much deeper sequencing.

**WITH A REFERENCE SEQUENCE**

A cost efficient solution to obtain molecular markers is to sequence DNA or RNA from the parental genotypes and mine polymorphisms from the resulting reads. These polymorphisms can be used later to design PCR markers or a high-throughput genotyping assay for the full population. This approach works remarkably well in diploid and polyploidy species using as low an amount of sequence as 5× coverage, meaning five times the size of the genome under study (Ossowski et al., 2008; Gore et al., 2009; Trick et al., 2009; Lai et al., 2010; Lam et al., 2010; Arai-Kichise et al., 2011; Gerald et al., 2011). A recent article reviews the methods and tools available for single nucleotide polymorphism (SNP) identification and genotyping (Nielsen et al., 2011). To align the reads to the reference, mapping softwares based in “seed methods” are preferred despite their slower nature because their robustness to polymorphisms. Before SNP calling users may consider removal of the reads that map to multiple locations in the reference, and of duplicated reads that may have been generated from PCR artifacts. A recent pipeline also recalibrates the quality of the nucleotides in the reads to correct for the high error rates in HTS, and realigns reads in complex genomic positions where the fast processing alignment algorithms may have failed (Depristo et al., 2011). Commonly used indicators of the veracity of polymorphisms are based in the amount and quality of reads showing the polymorphism, frequency of the observed alleles, quality of the alignment, and/or proximity to other polymorphisms. There are some basic and popular options for calling polymorphisms from aligned reads (Li et al., 2009a,b; Depristo et al., 2011), tools specialized in the analysis of reads from particular sequencing platforms (Souaiaia et al., 2011), that have the ability to detect structural variation (Chen et al., 2009; Hormozdiari et al., 2009, 2010), or that have into account the quality of the reference in addition to the quality of the reads (Frohler and Dieterich, 2010). An essential method to control for the quality of the data analysis process is visual inspection through genome viewers specialized in HTS datasets (Huang and Marth, 2008; Bao et al., 2009; Milne et al., 2010; Robinson et al., 2011).

**WITHOUT A REFERENCE SEQUENCE**

High-throughput sequencing sequences can serve to construct the necessary reference to identify molecular markers if it is not already available. Although assembling de novo a complete genome sequence is possible with HTS, it requires very deep sequencing and extensive bioinformatic analysis, even more given the relatively large size of most plant genomes. A more efficient option is sequencing mRNA, which greatly reduces sample complexity in comparison with genome sequencing and has the advantage of offering functional information such as coding polymorphisms or expression levels (Graham et al., 2010; Mizrachi et al., 2010; Bancroft et al., 2011; Everett et al., 2011; Garg et al., 2011; Guo et al., 2011; Ibarra-Laclette et al., 2011; Ness et al., 2011; Su et al., 2011; Wei et al., 2011). A comprehensive compilation of the methods and tools available for transcriptome assembly has been recently published (Martin and Wang, 2011). De novo assembly algorithms greatly benefit from long and paired-end reads, but are extremely sensitive to errors and polymorphisms and will not perform well during assembly of datasets from mixed genotypes or highly heterozygous individuals. The amount of new genomic positions detected in RNA-seq experiments decrease exponentially as the number of reads increases (Figure 2). The majority of medium and highly expressed transcripts in a sample are detected at low coverage, and increasing coverage will mainly add non-coding RNAs and low expressed transcripts at a very high cost (Tarazona et al., 2011). If the objective is to assemble complete transcriptomes, obtaining samples from diverse tissues, time points, and conditions is preferred to depth of sequencing. Even in the best possible conditions assemblies from RNA-seq reads will return only a subset of the existing transcripts, many of which will be fragmented. This is expected due to low expression of particular transcripts, the non-uniform read coverage, and the presence of different isoforms per gene. To help assembly of low expressed transcripts researchers can use normalization protocols that deplete the most abundant transcripts from the samples (Christodoulou et al., 2011). In any case, contigs resulting from de novo assembly can be effectively used as a reference for molecular marker detection and characterization of transcripts in un-sequenced genomes (Parchman et al., 2010; Wang et al., 2010; Angeloni et al., 2011; Hiremath et al., 2011; Kaur et al., 2011).

When highly similar genotypes are compared, RNA-seq may not be the best option since it mostly targets coding regions, which are less diverse than non-coding regions. In these cases researchers can construct reduced representation libraries by shearing DNA using restriction endonucleases and size-selecting the fragments that will be sequenced. Reads from these libraries can be clustered by similarity and mined for polymorphisms close to the restriction sites; or used to detect the presence–absence of particular tags, indicating a polymorphism in the restriction site itself (Kerstens et al., 2009; Sanchez et al., 2009; Etter et al., 2011). Obtaining polymorphisms from reduced representation libraries is more efficient when a reference sequence is available (Van Tassell et al., 2008; Wu et al., 2010). However, researchers have already developed tools to genotype samples from these tags using a low number of reads.

---

**FIGURE 2** | Percentage of transcriptome covered versus number of RNA-seq reads used. Eighty-one base pair paired-end RNA-seq reads from *S. lycopersicum* were randomly sampled in different subset sizes and aligned to the *S. lycopersicum* genome reference. The percentage of the length of the transcriptome covered by at least one read is represented at different coverages.
of reads from organisms without a reference (Ratan et al., 2010), or to reconstruct part of the targeted genome using paired-end sequencing (Willing et al., 2011). Additional protocols to obtain markers from reduced representation libraries exist in which different combination of restriction enzymes are used for each of the genotypes involved (Hyten et al., 2010), or that do not shear the DNA but filter the reads for single copy sequences (You et al., 2011). The amount of reads necessary to perform this type of analysis depends on the size of the genome, the restriction enzymes used, and the availability of a reference.

GENOTYPING POPULATIONS

With the price drop of the HTS technologies and the possibility of multiplexing samples, genotyping an entire population has become realistic (Schneebberger and Weigel, 2011). In the case of a sequenced system such as rice, generating reads from the individuals of a population at 0.02–0.055× coverage allowed high-density genotyping by comparisons with the parental genotypes (Huang et al., 2009), or by inferring the parental genotypes from the polymorphisms found in the population (Xie et al., 2010). Since erroneous polymorphism calls are expected at low coverage, more or less complex algorithms need to be defined to correctly genotype each polymorphism in each individual (Huang et al., 2009; Xie et al., 2010; Li et al., 2011). In addition, a reference sequence can serve researchers to design enrichment essays that will target their preferred genomic locations, although at high cost (Blow, 2009; Mamanova et al., 2010; Nijman et al., 2010; Kenny et al., 2011). For species where a genome sequence is not available, a very practical approach is to sequence reduced representation libraries as mentioned above (Baier et al., 2008; Emerson et al., 2010b; Hohenlohe et al., 2010, 2011).

MOLECULAR PHENOTYPING

The list of molecular phenotypes that can be quantified with HTS is extensive and is rapidly increasing (Hawkins et al., 2010). Examples of these phenotypes are protein–RNA interactions (Licatalosi et al., 2008; Hafner et al., 2010), translation rates (Ingolia et al., 2009; Ingolia, 2010), transcription rates (Core et al., 2008; Churchman and Weissman, 2011), protein–DNA interactions (Albert et al., 2007; Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007; Chen et al., 2008; Hesselberth et al., 2009), RNA degradation rates (Addo-Quaye et al., 2008; German et al., 2008), RNA secondary structure (Kertesz et al., 2008; German et al., 2008), natural antisense transcription (Cloonan et al., 2008; Core et al., 2008; He et al., 2008; Armour et al., 2008; Parkhomchuk et al., 2009) or small RNA profiles (Lu et al., 2005). QTL analysis using these phenotypes as traits is an exciting field that remains un-explored. Therefore, the computational frameworks to quantitatively compare these phenotypes between individuals will need to be established.

EXPRESSION PROFILING WITH HTS

Although many cases of phenotypic variation caused by coding polymorphisms have been documented, variation in gene expression has been shown to underlie much of phenotypic diversity (Reviewed in Hoekstra and Coyne, 2007; Wray, 2007; Stern and Orgogozo, 2008). One method to detect differences in expression between individuals using HTS is to sequence 26–27 nucleotide-long tags from expressed transcripts (Matsumura et al., 2010; Hong et al., 2011). A recent study shows that this method reaches saturation in mice with 6–8 million reads per sample (Hong et al., 2011). Its advantages over sequencing full transcripts are the lower cost, higher sensitivity, reduced bias during amplification due to the fixed fragment lengths, and use of simplified statistical models to calculate differential expression. On the other hand, methods based in tags will not detect the majority of coding polymorphisms and isoforms, and require a close enough reference sequence to extract biologically meaningful results.

RNA-seq is rapidly becoming a standard in expression profiling because of its simple protocol of preparation, digital nature, large dynamic range, and high sensitivity in comparison with previous technologies (Marioni et al., 2008; Bradford et al., 2010; Liu et al., 2010). In addition, it can serve to genotype individuals, identify novel transcripts, characterize alternative splicing, and quantify allele specific expression (Reviewed in Wang et al., 2009; Costa et al., 2010; Marguerat and Bahler, 2010). Due to the novelty of the technique there is no consensus on which sample preparation protocols present fewer biases (Raz et al., 2011). However, strand-specific methods could become a standard because of their increased precision due to their ability to distinguish between sense and antisense transcripts (He et al., 2008; Levin et al., 2010). In terms of experimental designs, it is necessary to randomize and replicate biological samples, with any other type of genome-wide analysis (Auer and Doerge, 2010; Fang and Cui, 2011; Hansen et al., 2011). There is little consensus about the depth of sequence needed for expression profiling with RNA-seq. Recent estimates range between 30 million reads to compare the expression profiles of two samples, to 100 million reads to detect most transcribed genes and quantify isoforms, to 500 million to obtain accurate profiles, including low expressed transcripts (Zhang et al., 2010; ENCODE, 2011; Toung et al., 2011). In any case, it is advisable to balance the number of reads between samples in the same experiment in order to perform accurate expression comparisons (Tarazona et al., 2011).

Expression profiling from HTS datasets is necessarily based on counting the reads mapped to each transcript in a reference sequence. When a reference genome or transcriptome is not available, it can be reconstructed using de novo assembly of the reads for at least one of the genotypes as described above. The simpler and less computational intensive protocol for expression profiling is to map the RNA-seq reads to known (or de novo assembled) transcripts and a set of possible exon–exon junctions (when available) to detect alternative splicing. However, in organisms with sequenced genomes this protocol will not allow detection of novel exons, transcripts, and isoforms. The preferred pipeline involves aligning the reads to the genomic reference using an alignment tool that splices the reads to detect intron–exon junctions (For example Trapnell et al., 2009; Ameur et al., 2010; Au et al., 2010; Guttman et al., 2010; Wang et al., 2010b; Lou et al., 2011).

A challenge for expression analyses in samples from two unrelated individuals is the need to perform robust quantification of reads generated from two or more alleles. This implies that reads
with the closer genotype to the reference will align better than those based in the Burrows–Wheeler Transform algorithm (For a review see Garber et al., 2011). Although most studies ignore this problem, there are solutions that go from identifying and removing the polymorphisms that cause these biases (Degner et al., 2009), aligning the reads to all references from the genotypes involved (Bullard et al., 2010a) or including the polymorphisms found in the references (Gan et al., 2011). When two references are used, a potential problem may arise from motifs that are more abundant in one reference with respect to the other if only uniquely mapped reads are counted. The use of longer reads and/or paired-end reads greatly decreases the number of ambiguously mapped reads. In addition, there are robust methods to assign this multi-mapped reads to a single location (Faulkner et al., 2008; Mortazavi et al., 2008; Hashimoto et al., 2009; Li et al., 2010a; Wang et al., 2010a; Ji et al., 2011).

There are a number of tools to count the number of reads aligned to each transcriptional unit to calculate expression, most of which require knowledge of Perl, Python, Linux/Unix, or R (Carlson et al., 2009; Bio:DB:Sam, 2009; Anders, 2010; Morgan and Pagès, 2010; Quinlan and Hall, 2010). Some alignment tools can directly calculate the number of reads per transcript and/or a measure of expression based in the reads (or fragments) per gene size in kilobases per million reads mapped, called RPKM (or FPKM; Mortazavi et al., 2008; Trapnell et al., 2010). However, these expression units show biases depending on the length, number, abundance of the transcripts present in the samples, or because of technical replication (Oshlack and Wakefield, 2009; Bullard et al., 2010b; McIntyre et al., 2011). For this reason researchers have developed dedicated R/Bioconductor packages to calculate differential expression between samples based on raw read counts per transcript (Anders and Huber, 2010; Bullard et al., 2010b; Hardcastle and Kelly, 2010; Robinson et al., 2010; Wang et al., 2010c). In addition, there are software packages that take into consideration the biases inherent to RNA-seq when calculating expression or performing downstream analyses such as gene ontology over-representation studies (Young et al., 2010; Zheng et al., 2011).

High-throughput sequencing datasets allow quantification of expression for each isoform separately, resulting in significantly more accurate estimates than calculating expression at the gene level (Wang et al., 2010d). For this, users must first identify splicing events from the reads that align to exon–exon junctions. Quantifying isoform expression is complicated since most reads in an alternatively spliced transcript cannot be assigned to a single isoform. The most promising methods to address this complex problem take advantage from the information offered by paired-end and/or unambiguously mapped reads (Guttman et al., 2010; Katz et al., 2010; Li et al., 2010a; Trapnell et al., 2010; Nicolae et al., 2011). One advantage of going through the intricate process of identification of alternative splicing is that it can also be used as a trait for QTL analysis (Li et al., 2010c; Montgomery et al., 2010; Pickrell et al., 2010; Lalonde et al., 2011).

### ALLELE SPECIFIC EXPRESSION IN HYBRIDS

An alternative to sequencing a full segregating population to perform eQTL analyses is to sequence F1 hybrid individuals, where allele specific expression can be calculated for loci with coding polymorphisms (Babak et al., 2008, 2010; Bullard et al., 2010a; Emerson et al., 2010a; Mcmanus et al., 2010; Pickrell et al., 2010). For any gene, both alleles in the hybrid share the same cellular environment and, as a result, changes in expression between alleles must necessarily be due to cis-acting regulators (Cowles et al., 2002). Trans-acting eQTLs can be inferred by performing RNA-seq in the parents and comparing the differences in expression levels between alleles in the hybrid with the differences between the parents (Wittkopp et al., 2004). Despite the considerable reduction in price and simplicity of experimental design, this method has several drawbacks. Allele specific expression can only be calculated in transcripts with coding polymorphisms that are highly covered, and it is very dependent on read and transcript length (Degner et al., 2009; Fontanillas et al., 2010). New statistical approaches are being developed that will overcome these limitations, starting by being able to estimate false discovery rates and allele specific alternative splicing (Skelly et al., 2011).

In summary, HTS is changing the way we perform QTL analysis by allowing high-throughput genotyping of populations and phenotyping of traits with a precision not achievable before. It is clear that HTS has not reached its peak of development, and that tools and algorithms will have to be modified according to the new technological improvements. Nevertheless, the first experiments using this technology have already identified exciting possibilities for the characterization of natural variation in plants.

### REFERENCES

Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr. Biol. 18, 758–762.

Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12, R18.

Albert, I., Mavrich, T. N., Tomsho, L. P., Qj, J., Zanton, S. J., Schuster, S. C., and Pugh, B. F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature 446, 572–576.

Amur, A., Wetterbom, A., Feuk, L., and Gyllensten, U. (2010). Global and unbiased detection of splice junctions from RNA-seq data. Genome Biol. 11, R34.

Anders, S. (2010). HTSeq: Analyzing High-Throughput Sequencing Data with Python. Available at: http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html#author

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Angeloni, E., Wagemaker, C. A., Jetten, M. S., Op Den Camp, H. J., Jansen-Megens, E. M., Francois, K. J., Stunnenberg, H. G., and Ouborg, N. J. (2011). De novo transcriptome characterization and development of genomic tools for Scabiosa columbaria L. using next-generation sequencing techniques. Mol. Ecol. Resour. 11, 662–674.

Arai-Kichise, Y., Shiwa, Y., Nagasaki, H., Ebana, K., Yoshikawa, H., Yano, M., and Wakasa, K. (2011). Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. Plant Cell Physiol. 52, 274–282.

Armour, C. D., Castle, J. C., Chen, R., Babak, T., Loehr, P., Jackson, S., Shah, J. K., Dey, J., Rohl, C. A.,
profiling of histone methylations in the human genome. Cell 129, 823–837.
Bioinformatics. (2009). Available at: search.ensembl.org/~ads/Bio-
SamTools/ibs/Bio/DB/Bam/ Alignment.pm
Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., and Nekrutenko, A. (2010). Manipula-
tion of FASTQ data with galaxy. Bioinformat-
sics 26, 1785–1785.
Blow, N. (2009). Genomics catch me if you can. Nat. Methods 6, 539–543.
Boyle, A. P., Davis, S., Shullia, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., and Crawford, G. E. (2008). High-resolution mapping and characterization of open chro-
matin across the genome. Cell 132, 741–755.
Bradford, J. R., Hey, Y., Yates, T., Li, Y., Pepper, S. D., and Miller, C. J. (2010). A comparison of mas-
ively parallel nucleotide sequenc-
ing with oligonucleotide microar-
rays for global transcription pro-
figling. BMC Genomics 11, 282.
doi:10.1186/1471-2164-11-282
Bullard, J. H., Mostovoy, Y., Dudoit, S., and Brem, B. R. (2010a). Polyno-
goic and directional regulatory evolution across pathways in Saccharomyces. Proc. Natl. Acad. Sci. U.S.A. 107, 5038–5063.
Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010b). Evaluation of statistical methods for normaliza-
tion and differential expression in mRNA-seq experi-
ments. BMC Bioinformatics 11, 94. doi:10.1186/1471-2105-11-94
Carlson, M., Pages, H., Aboyoun, P., Falcon, S., Morgan, M., Sarkar, D., and Lawrence, M. (2009). Genomic
Features: Tools for Making and Mani-
pulating Transcript Centric Anno-
tations. Available at: http://www.
bioconductor.org/packages/2.6/bioc/html/GenomicFeatures.html
Chen, K., Wallis, J., Mcclellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., Mcgrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L., and Mardis, E. R. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods 6, 677–681.
Chen, X., Xu, H., Yuan, P., Fang, E., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W. K., Clarke, N. D., Wang, J., Liu, Y., and Shi, S. (2009). MapView: visualization of short reads align-
ment on a desktop computer. Bioin-
formatics 25, 1554–1555.
Bao, S., Jiang, R., Kwan, W., Wang, B., Ma, X., and Song, Y. Q. (2011). Evaluation of next-generation sequencing software in mapping and assembly. J. Hum. Genet. 56, 406–414.
Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution
Everett, M. V., Grau, E. D., and Seeb, J. E. (2011). Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. Mol. Ecol. Resour. 11(Suppl. 1), 93–108.

Falgueras, J., Lara, A. J., Fernandez-Pozo, N., Canton, F. R., Perez-Trabado, G., and Claros, M. G. (2010). SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. BMC Bioinformatics 11, 38. doi:10.1186/1471-2105-11-38

Fang, Z., and Cui, X. (2011). Design and validation issues in RNA-seq experiments. Brief. Bioinform. 12, 280–287.

FastQC. (2008). Available at: http://www.bioinformatics.bbsrc.ac.uk/ projects/fastq/

FASTX-Toolkit. (2009). Available at: http://hannonlab.cshl.edu/fastx-toolkit/index.html

Falgueras, J., Lara, A. J., Fernandez-Pozo, A. J., Fan, L., Koziol, M. J., Grinker, A., Nusbaum, C., Rinn, J. L., Lander, E. S., and Regev, A. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. 28, 503–510.

Hafner, M., Landthaler, M., Burger, L., Khoshrud, M., Hauser, J., Berninger, P., Rothbächer, A., Ascano, M. J., Ziegler, S., Friedmann, M., Birrol, L., Jones, S. J., Cronk, Q. C., and Douglas, C. J. (2011). SNP discovery in black cottonwood (Populus trichocarpa) by population transcriptome resequencing. Mol. Ecol. Resour. 11(Suppl. 1), 81–92.

German, M. A., Pilla, J., Jeong, D. H., Bohnert, R., Jean, G., Derwent, P. R., Leisch, F., Li, C., Maechler, M., Rinn, J. L., Lander, E. S., and Buckler, E. S. (2009). A first-generation whole genome assembly for Arabidopsis thaliana. Nature 458, 33–36.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 11, R86.

Hansen, K. D., Brenner, S. E., and Dudoit, S. (2010). Bias in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 38, e131.

Hansen, K. D., Zhijin, W., Irizarry, R. A., and Leek, J. T. (2011). Sequencing technology does not eliminate biological variability. Nat. Biotechnol. 29, 575–573.

Hardcastle, T. J., and Kelly, K. A. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11, 422. doi:10.1186/1471-2105-11-422

Hawkins, R. D., Hon, G. C., and Ren, B. (2009). Identification of transcription start sites by P AR-CLIP. Cell 137, 335–347.

Huang, W., and Martha, G. (2008). EagleView: a genome assembly viewer for next-generation sequencing technologies. Genome Res. 18, 1538–1543.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T., and Han, B. (2009). High-throughput genotyping by whole-genome resequencing. Genome Res. 19, 1068–1076.
Kerstens, H., Cronjöven, R., Veena-{
}daal, A., Dibbins, B., Chinn-a-Woeng, T., Den Dunnen, J., and Groe-
{n}ne, M. (2009). Large scale single nucleotide polymorphism discov-
{r}y in unsequenced genomes using second generation high through-
{p}ut sequencing technology: applied to turkey. BMC Genomics 10, 479.
{d}oi:10.1186/1471-2164-10-479

Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y.,
{a}nd Segal, E. (2010). Genome-wide measurement of RNA secondary
structure in yeast. Nature 467, 103–107.

Keurentjes, J. J. B., Fu, J., Terpstra, I. R. G., Garcia, J. M., Van Den Ack-
erveken, G., Snoek, L. B., Peeters, A. J. M., Vreugdenhil, D., Koorn-
manning, M., and Jansen, R. C. (2007). Regulatory network construction in
Arabidopsis by using genome-wide gene expression quantitative trait loci.
Proc. Natl. Acad. Sci. U.S.A. 104, 1708–1713.

Kirst, M., Basten, C. J., Myburg, A. A., Zeng, Z. B., and Sederoff, R. R.
(2005). Genetic architecture of transcript-level variation in different-
iating xylem of a eucalyptus hybrid. Genetics 169, 2295–2303.

Kliebenstein, D. J., West, M. A., Van Leeuwen, H., Luedt, O., Doerge, R.
W., and St Clair, D. A. (2006). Identification of QTLs controlling gene
expression networks defined as a priori. BMC Bioinformatics 7, 308.
{d}oi:10.1186/1471-2105-7-308

Lai, L.-J., R., Xu, X., Jin, W., Xu, M.-Z., H. B., Xiang, Z., Song, W., Ying, K.
Zhang, M., Jiao, Y., Ni, P., Zhang, J., Li, D., Guo, X., Ye, K., Jian, M., Wang,
B., Zheng, H., Liang, H., Zhang, X., Wang, S., Chen, S., Li, J., Fu, Y.,
Springer, N. M., Yang, H., Wang, J., Dai, Y., Rojas, C. C., and Deng, H.
(2011). Genome-wide patterns of genetic variation among elite maize inbred
lines. Nat. Genet. 42, 1027–1030.

Lalonde, E., Ha, K.-C., Wang, Z., Bemmo, A., Kleinman, C. L., Kwan, T., Pasti-
en, T., and Majewski, J. (2011). RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome
Research 21, 545–554.

Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., Li, M. W., He,
W., Qin, N., Wang, B., Li, J., Jian, M., Wang, J., Shao, G., Sun, S. S.,
and Zhang, G. (2010). Resequencing of 31 wild and cultivated soy-
bean genomes identifies patterns of genetic diversity and selection. Nat.
Genet. 42, 1053–1059.

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedmann, N., Gnikre, A., and
Regev, A. (2010). Comprehensive comparative analysis of strand-
specific RNA sequencing methods. Nat. Methods 7, 799–715.

Li, B., Ruotti, V., Stewart, R. M., Thom-
son, J. A., and Dewey, C. N. (2010a).
RNA-seq gene expression estimation with read mapping uncertainty.
Bioinformatics 26, 493–500.

Li, J., Jiang, H., and Wong, W. H.
(2010b). Modeling non-uniformity in short-read rates in RNA-seq data.
Genome Biol. 11, R50.

Li, Y., Breiling, S., Snoek, L. B., Van Der Velde, K. J., Swertz, M. A., Riksen,
J. K., Crooijmans, R., Van Den Ack-
ervoet, P. M., and Forster, J. W. (2010).
BAMmapper: a suffix array-based
approach to construct genotype dis-
coveries. Bioinformatics 27, 421–422.

Lu, C., Tei, S. S., Luo, S., Haudenschild,
C. D., Meyers, B. C., and Green, P. J. (2005). Elucidation of the small
RNA component of the transcripti-
some. Science 309, 1567–1569.

Mamanova, L., Coeyffy, A. J., Scott,
C. E., Kozarewa, I., Turner, E. H.,
Kumar, A., Howard, E., Shendure, J.,
and Turner, D. J. (2010). Target-
engaged antibody discovery for next
 generation sequencing. Nat. Meth-
ods 7, 111–118.

Marguerat, S., and Bahler, J. (2010).
RNA-seq from technology to biology.
Cell. Mol. Life Sci. 67, 569–579.

Marioni, J. C., Mason, C. E., Mane, S. M.,
Stephens, M., and Gilad, Y. (2008).
RNA-seq an assessment of techni-
cal reproducibility and comparison
with gene expression arrays. Genome
Biol. 18, 1509–1517.

Martin, J. A., and Wang, Z. (2011).
Next-generation transcriptome as-
sembly. Nat. Rev. Genet. 12, 671–682.

Matzmann, H., Yoshida, K., Luo, S.,
Kumira, E., Fujibie, T., Albertzen, Y.,
Burrero, R. A., Kruger, D. H., Kahl, G.,
Schroth, G. P., and Terauchi, R.
(2010). High-throughput Super-
SAGE for digital gene expression analysis of multiple samples using
next generation sequencing. PLoS ONE 5, e12010. doi:10.1371/jour-
nal.pone.0012010

McIntyre, L. M., Lopiano, K. K., Morse,
A. M., Amin, V., Oberg, A. L., Young,
L. J., and Nuzhdin, S. V. (2011).
RNA-seq technical variability and sampling. BMC Genomics 12, 293.
{d}oi:10.1186/1471-2164-12-293

Meinmann, C. J., Coolon, J. D., Duff,
M. O., Epper-Mains, J., Graveley, B. R., and Wiltkopp, P. J. (2010).
Regulatory divergence in Drosophila revealed by RNA-seq. Genome
Res. 20, 816–825.

Meisner, A., Mikkelsen, T. S., Gu., Huang,
Wernig, M., Hanna, J., Svachekno,
A., Zhang, X., Bernstein, B. E.,
Nusbaum, C., Jaffe, D. B., Gnikre,
A., Jaenisch, R., and Lander, E. S.
(2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766–770.

Metzker, M. L. (2010). Sequencing tech-
nologies – the next generation. Nat.
Rev. Genet. 11, 31–46.

Frontiers in Plant Science | Plant Physiology

November 2011 | Volume 2 | Article 77 | 8
Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Gnansou, G., Alvarez, P., Brockman, W., Kim, T. K., Kohce, R. P., Lee, W., Mendenhall, E., O’donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E. S., and Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, E., and Marshall, D. (2010). Tablet – next generation sequence assembly visualization. Bioinformatics 26, 401–402.

Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F., and Myburg, A. A. (2010). De novo assembly in cross- and self-fertilizing plants. Third-generation sequencing fire-works at Marco Island. Nat. Biotechnol. 28, 426–428.

Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbet, J., Guigo, R., and Dermitzakis, E. T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773–777.

Morgan, M., and Pagis, H. (2010). Rsamtools Import Aligned BAM File Format Sequences Into R/Bioconductor. Available at: http://bioconductor.org/packages/release/bioc/html/Rsamtools.html

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes using RNA-seq. Nat. Methods 5, 621–628.

Munroe, D. J., and Harris, T. J. (2010). Third-generation sequencing fire-works at Marco Island. Nat. Biotechnol. 28, 426–428.

Ness, R. W., Siol, M., and Barrett, S. C. (2011). De novo sequence assembly and characterization of the fl oral transcriptome in cross- and self-fertilizing plants. BMC Genomics 12, 298. doi:10.1186/1471-2164-12-298

Nicolae, M., Murgulescu, A., Mander, M., and Zelikovsky, A. (2011). Estimation of alternative splicing isoform frequencies from RNA-seq data. Algorithms Mol. Biol. 6, 9.

Niedringhaus, T. P., Milanovski, D., Kerby, M. B., Snyder, M. P., and Barron, A. E. (2011). Landscape of next-generation sequencing technologies. Anal. Chem. 83, 4327–4341.

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12, 443–451.

Nijman, I. J., Mokry, M., Van Boxtel, R., Toonen, P., De Bruijn, E., and Cuppen, E. (2010). Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. Nat. Methods 7, 913–915.

Oshlack, A., and Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. Biol. Direct 4, 14.

Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. 18, 2074–2083.

Parchomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lebrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 37, e123.

Pickrell, J. K., Marioni, J. C., Pai, A., Degner, J. F., Engelhardt, B. E., Nkadori, E., Veyrieras, J. B., Stephens, M., Gilad, Y., and Richardson, P. M. (2008). Rapid human gene expression variation with short reads. Nat. Biotechnol. 26, 426–428.

Plessey, C., Bertin, N., Takahashi, H., Simone, R., Salimullah, M., Lassmann, T., Vitezce, M., Severin, J., Oliviari, S., Lazarevic, D., Hornig, N., Orlando, V., Bell, I., Gao, H., Dumais, J., Kaplanov, P., Wang, H., Davis, C. A., Gingeras, T. R., Kawai, J., Daub, C. O., Hayashizaki, Y., Gustinich, S., and Carninci, P. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772.

Pocock, M. R., Schattner, P., Senger, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Ruffalo, M., Laframboise, T., and Koyturk, M. (2011). Comparative analysis of algorithms for next-generation sequencing read alignment. Bioinformatics 27, 2790–2796.

Sanchez, C. C., Smith, T. P., Wiedmann, R. T., Vallejo, R. L., Salem, M., Yao, J., and Rexroad, C. E. III. (2009). Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. BMC Genomics 10, 559. doi:10.1186/1471-2164-10-559

Schlotterer, C. (2004). The evolution of molecular markers – just a matter of fashion? Nat. Rev. Genet. 5, 63–69.

Schneider, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.

Schneider, R., Lim, Y. W., Rohwer, F., and Edwards, R. (2010). TagCleaner: identification and removal of tag sequences from genomic and metagenomic datasets. BMC Bioinformatics 11, 341. doi:10.1186/1471-2105-11-341

Schweberer, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., Jorgensen, J. E., Weigel, D., and Andersen, S. U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods 6, 550–551.

Schweberer, K., and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci. 16, 282–288.

Schwartz, S., Oren, R., and Ast, G. (2011). Detection and removal of biases in the analysis of next-generation sequencing reads. PLoS ONE 6, e16665. doi:10.1371/journal.pone.0016685

Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J., and Akey, J. M. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res. doi:10.1101/gr.119784.110. [Epub ahead of print].

Smith, D. R., Quinlan, A. R., Peckham, H. E., Makowsky, K., Tao, W., Woolf, B., Shen, L., Donahue, W. F., Tusneem, N., Stromberg, P. M., Swart, A. D., Zhang, L., Ranade, S. S., Warner, J. B., Lee, C. C., Coleman, B. E., Zhang, Z., Mclaughlin, S. F., Malek, J. A., Sorensen, J. M., Blanche, A. P., Chapman, J., Hillman, D., Chen, F., Rokhsar, D. S., Mckenan, J. N., Jeffries, T. W., Marti, G. T., and Richardson, P. M. (2008). Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res. 18, 1638–1642.

Souzaia, T., Fraizer, Z., and Chen, T. (2011). Combl: SNP calling and mapping analysis for color and nucleotide space platforms. J. Comput. Biol. 18, 795–807.

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdijian, C., Fuellen, G., Gilbert, J. G., Korl, I., Lapp, H., Levinas, H., Matsalla, C., Mungall, C. J., Osborne, B. L., Pocock, M. R., Schattner, P., Serger, M., Stein, L. D., Stupka, E., Wilkinson, M. D., and Birney, E. (2002). The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12, 1611–1618.

Stein, L. D. (2010). The case for cloud computing in genome informatics. Genome Biol. 11, 207.

Stern, D. L., and Onggogozu, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution 62, 2155–2177.
Su, C. L., Chao, Y. T., Alex Chang, Y. C., Chen, W. C., Chen, C. Y., Lee, A. Y., Hwa, K. T., and Shih, M. C. (2011). De novo assembly of expressed transcripts and global analysis of the Phalaenopsis aphrodite transcriptome. Plant Cell Physiol. 52, 1501–1514.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome Res. doi: 10.1101/gr.124321.111. [Epub ahead of print].

Young, J. M., Morley, M., Li, M., and Cheung, V. G. (2011). RNA-sequence analysis of human B-cells. Genome Res. 21, 991–998.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25, 1105–1111.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515.

Trick, M., Long, Y., Meng, J., and Ban- croft, I. (2009). Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnol. 17, 334–346.

Underwood, J. G., Uzokwu, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., Lowe, T. M., Salama, S. R., and Haussler, D. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat. Methods 7, 995–1001.

Van Tassell, C. P., Smith, T. P., Matuku-malli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., Haudenschild, C. D., Moore, S. S., Warren, W. C., and Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Methods 5, 247–252.

Vulsteke, M., Delee, H., Vercauteren, A., Zabeau, M., and Kuiper, M. (2006). Genetic dissection of transcriptional regulation by cDNA-AFLP. Plant J. 43, 439–446.

Vulsteke, M., Van Eeuwijk, F., Van Hummelen, P., Kuiper, M., and Zabeau, M. (2005). Genetic analysis of variation in gene expression in Arabidopsis thaliana. Genetics 171, 1267–1275.

Wang, J., Huda, A., Lunyak, V. V., and Jordan, I. K. (2010a). A Gibbs sampling strategy applied to the mapping of ambiguous short-sequence tags. Bioinformatics 26, 2501–2508.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., He, X., Mieczkowski, P., Grimm, S. A., Perou, C. M., Macleod, J. N., Chiang, D. Y., Prins, J. F., and Liu, J. (2010b). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 38, e178.

Wang, L., Feng, Z., Wang, X., and Zhang, X. (2010c). Diff Seq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136–138.

Wang, X., Wu, Z., and Zhang, X. (2010d). Isoform abundance inference provides a more accurate estimation of gene expression levels in RNA-seq. J. Bioinform. Comput. Biol. 8(Suppl. 1), 177–192.

Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., Chen, X., and Li, Y. (2010e). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (Ipomoea batatas). BMC Genomics 11, 726. doi:10.1186/1471-2164-11-726

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–65.

Wei, W., Qi, X., Wang, L., Zhang, Y., Hua, W., Li, D., Iv, H., and Zhang, X. (2011). Characterization of the sesame (Sesamum indicum L) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics 12, 451. doi:10.1186/1471-2164-12-451

West, M. A. L., Kim, K., Kliebenstein, D. J., Van Leeuwen, H., Michelmore, R. W., Doerge, R. W., and St Clair, D. A. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175, 1441–1450.

Willing, E. M., Hoffmann, M., Klein, J. D., Weigel, D., and Dreyer, C. (2011). Paired-end RAD-seq for de novo assembly and marker design without available reference. Bioinformatics 27, 2187–2193.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. Nature 430, 85–88.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8, 206–216.

Wu, X., Ren, C., Joshi, T., Vuong, T., Xu, D., and Nguyen, H. T. (2010). SNP discovery by high-throughput sequencing in soybean. BMC Genomics 11, 469. doi:10.1186/1471-2164-11-469

Xie, W., Feng, Q., Yu, H., Huang, X., Zhao, Q., Xing, Y., Yu, S., Han, B., and Zhang, Q. (2010). Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc. Natl. Acad. Sci. U.S.A. 107, 10578–10583.

You, F. M., Huo, N., Deal, K. R., Gu, Y. Q., Luo, M. C., Mgueur, P. E., Dvorak, J., and Anderson, O. D. (2011). Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. BMC Genomics 12, 59. doi:10.1186/1471-2164-12-59

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11, R14.

Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhang, R., Lu, Z., He, Z., Fang, X., Chen, L., Tian, W., Tao, Y., Kristiansen, K., Zhang, X., Li, S., Yang, H., and Wang, J. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res. 20, 646–654.

Zheng, W., Chung, L. M., and Zhao, H. (2011). Bias detection and correction in RNA-sequencing data. BMC Bioinformatics 12, 290. doi:10.1186/1471-2105-12-290

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 April 2011; accepted: 23 October 2011; published online 15 November 2011.

Citation: Jiménez-Gómez JM (2011) Next generation quantitative genetics in plants. Front. Plant Sci. 2:77. doi: 10.3389/fpls.2011.00077

This article was submitted to Frontiers in Plant Physiology, a specialty of Frontiers in Plant Science.

Copyright © 2011 Jiménez-Gómez. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.