Agricultural Reviews, 38(4) 2017 : 271-281
Print ISSN:0253-1496 / Online ISSN:0976-0539

AGRICULTURAL RESEARCH COMMUNICATION CENTRE
www.arccjournals.com

Transcriptome profiling: methods and applications- A review

Bibha Rani* and V.K. Sharma

Rajendra Agricultural University, Pusa,
Samastipur-848 125, Bihar, India.

Received: 26-06-2016 Accepted: 19-09-2017 DOI: 10.18805/ag.R-1549

ABSTRACT

Global transcriptional profiling is a powerful tool that can expose expression patterns to define cellular states or to identify genes with similar expression patterns. In recent years, transcriptome profiling has been widely used to understand the genetic regulation of a particular cell type. Transcriptome is defined as a full range of messenger Ribonucleic acid (RNA) molecule expressed by an organism. In other words a transcriptome represents the small percentage of genetic code that is transcribed into RNA molecules. It can offer valuable information on the significant biological processes behind the maintenance of the functionality of the cell. Transcriptomics provides fundaments for more definitively designed studies and guidance to select the genes for functional studies. The technology for the study of the transcriptome is not dependent on any prior knowledge of the genes expressed in the cells. However, with regards to the administration and interpretation of the enormous data provided by transcriptome profiling challenges remain.. Four methods have been reviewed here that is, Microarray technology, Serial Analysis of Gene Expression (SAGE), RNA sequencing (RNA-Seq) and Massively Parallel Signature Sequencing (MPSS). The use of these technologies to analyse the expressed transcripts in several prokaryotic and eukaryotic genomes has revealed the high complexity of transcriptomes.

Key words: Microarray, MPSS, Sage.

The Genome is a store of biological information but on its own, it is unable to release that information to the cell. Utilization of the biological information requires the coordinated activity of enzymes and other proteins, which participate in a complex series of biochemical reactions referred to as genome expression. The initial product of genome expression is the transcriptome, the entire repertoire of transcripts in a species, represents a key link between DNA and phenotype whose biological information is required by the cell at a particular time. These RNA molecules direct synthesis of the final product of genome expression, the proteome, the cell's repertoire of proteins, which specifies the nature of the biochemical reactions that the cell is able to carry out. Gene expression profiles can be obtained and compared by various methods, such as RNA-DNA hybridization measurements (Harrington *et.al*., 2000), subtractive hybridization (Byers *et.al*., 2002), subtraction cDNA libraries (Jiang *et.al*., 2002), and differential display (Lievens *et.al* 2001). However, these methods have been limited in providing overall gene expression patterns due to their technical shortcoming.

Several high throughput methods of transcriptome profiling have been developed with two basic approaches, hybridization-based method (Microarray technology) and sequencing based methods (RNA sequencing, MPSS, SAGE), both offering great opportunities for large scale analysis.

PRE-mRNA PROCESSING AND ALTERNATE SPLICING: The discovery that gene sequences are interrupted by noncoding segments (introns) that are removed during message processing (Berget *et. al*., 1977) was initially surprising, but mRNA processing is now known to be common in eukaryotic genes. Most intron splicing is carried out by the spliceosome, a large macromolecular machine composed of five small nuclear riboproteins (snRNPs) and numerous accessory proteins (Staley and Guthrie, 1998). In metazoans, intron removal and the joining of flanking exons is directed by four sequence signals: the exon–intron junctions at the 5' end and 3' end that are the splice donor and acceptor sites, respectively, and two sites within the introns-the branch site sequence located upstream of the 3' splice site, and the polypyrimidine tract located between the 3' splice site and the branch site. Interestingly, in plants the pyrimidine tracts are mostly uridine, and the branch point sequences are not obvious (Reddy, 2007). Although plant genomes are known to encode homologs of many proteins that are included in animal spliceosomes, plant spliceosomes have never been isolated, and their exact protein composition is yet unverified.

Alternative splicing (AS) creates multiple mRNA transcripts, or isoforms, from a single gene (Fig.1). While AS had been observed in several genes by the early 1980s (Early *et al*., 1980; Rosenfeld *et al*., 1982), it was

*Corresponding author's e-mail: bibha9rani@gmail.com.

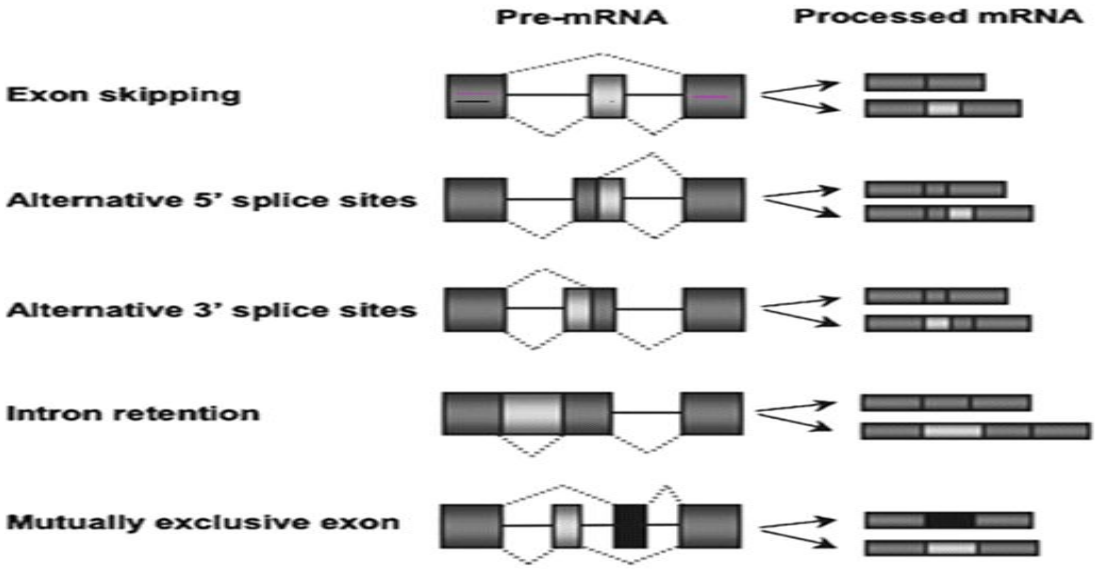


Fig-1: Common types of alternative splice events.

characterized at the single gene level and thought to occur in <5% of human genes (Sharp, 1994). However, analysis of genome sequence data has demonstrated that AS is widespread in metazoans (Sorek and Ast, 2003; Kim *et al*., 2007). While AS in humans is known to be common, AS in plants was not extensively observed and previously thought to be rare (Brett *et. al*., 2002). Recent computational and experimental studies suggest that alternative splicing plays a far more significant role in the generation of proteome diversity in plants than previously thought (Xing and Lee, 2006).

Microarray technology: The most commonly used technology to profile the expression of thousands of transcripts simultaneously are microarrays. cDNA and oligonucleotide arrays are two types of platforms commonly used. In cDNA arrays, cDNAs from a clone collection or cDNA library are spotted on nylon membrane or glass slide (Fig.2). As many as 30,000 fragments can be spotted on a microscope slide with each spot corresponding to a unique cDNA (Eisen *et.al*., 1998). The second type of microarray uses oligonucleotides. These are either etched on a silicon chip by photolithography or are printed on glass slides using ink jet technology. The oligonucleotide or cDNA spotted array is hybridized to cDNAs synthesized from the mRNA or total RNA extracted from the cell or tissue of interest. The cDNA from two different samples are labeled with fluorescent dyes such as Cy3 (green) and Cy5 (red). These samples can be different cell populations or treatment conditions. The cDNA labeled with Cy3 and Cy5 are mixed together and hybridized against the same array. The two populations compete for the same targets or probe spots on the array (Fig 3). The array is scanned with two different wavelengths following hybridization and washing. The spot intensity at the two wavelengths is determined. A ratio or log ratio between the two fluorescent intensities is calculated (Danila *et. al*., 2010).

Result analysis with array mining: Array Mining.net is a web-application for microarray analysis that provides easy access to a wide choice of feature selection, clustering, prediction; gene set analysis and cross-study normalization methods (Table 1). The most common task in statistical microarray analysis is gene selection, sample clustering, sample classification and gene set analysis (Table 2).

Serial analysis of gene expression (SAGE): SAGE is a sequence-based approach which was first introduced in 1995 by Velculescu and coworkers. It allows identification of a large number of transcripts present in tissues and the quantitative comparison of transcriptomes. The method is based on generation of a short specific tag (14 bp) from each mRNA present in a sample, resulting in the production of a SAGE tags library representative of this sample. The sequencing of these tags allows a high-throughput determination of their frequencies in the library, which are correlated with the relative amounts of the corresponding mRNAs. Thus, thousands of different transcripts can be analyzed, with a high specificity and most importantly, without any a prior knowledge of their identity. SAGE has proven to be a very powerful and robust method for investigating gene expression at the whole-genome scale (Boon *et.al*., 2002) and to reflect the actual relative contents of mRNAs in a sample. As compared with cDNA arrays or oligochips, it has several advantages, such as the possibility to perform transcript profiling without the need of large technological investments and the ability to obtain comprehensive transcriptomes from minute amounts of RNA (Virlon *et. al*., 1999). The SAGE technology has been used extensively with animal systems, and more particularly in cancer research, where several hundred libraries and nearly

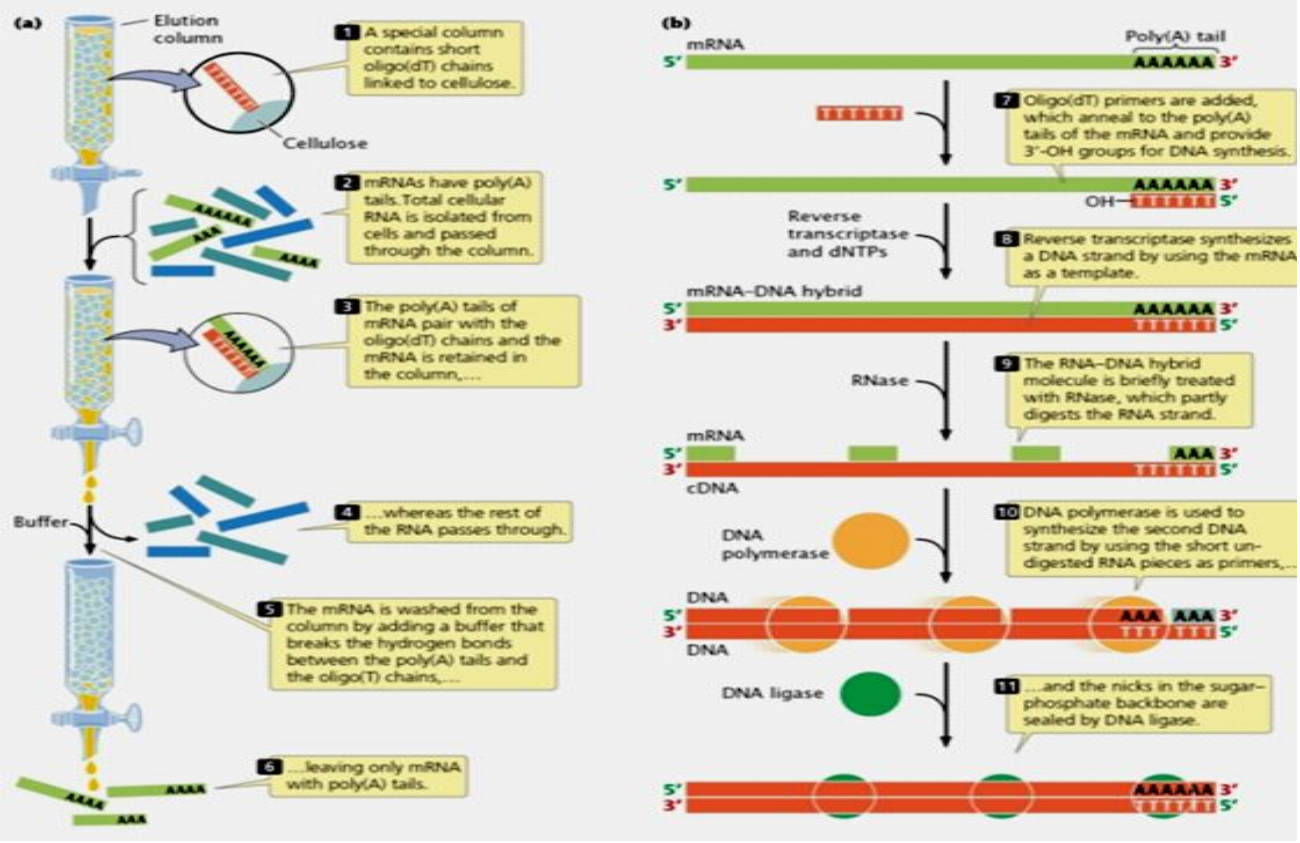


Fig-2: (Courtesy: W. H. Freeman Pierce, Benjamin. *Genetics 2005: A Conceptual Approach*, 2nd ed.) Approaches to construction of cDNA libraries.

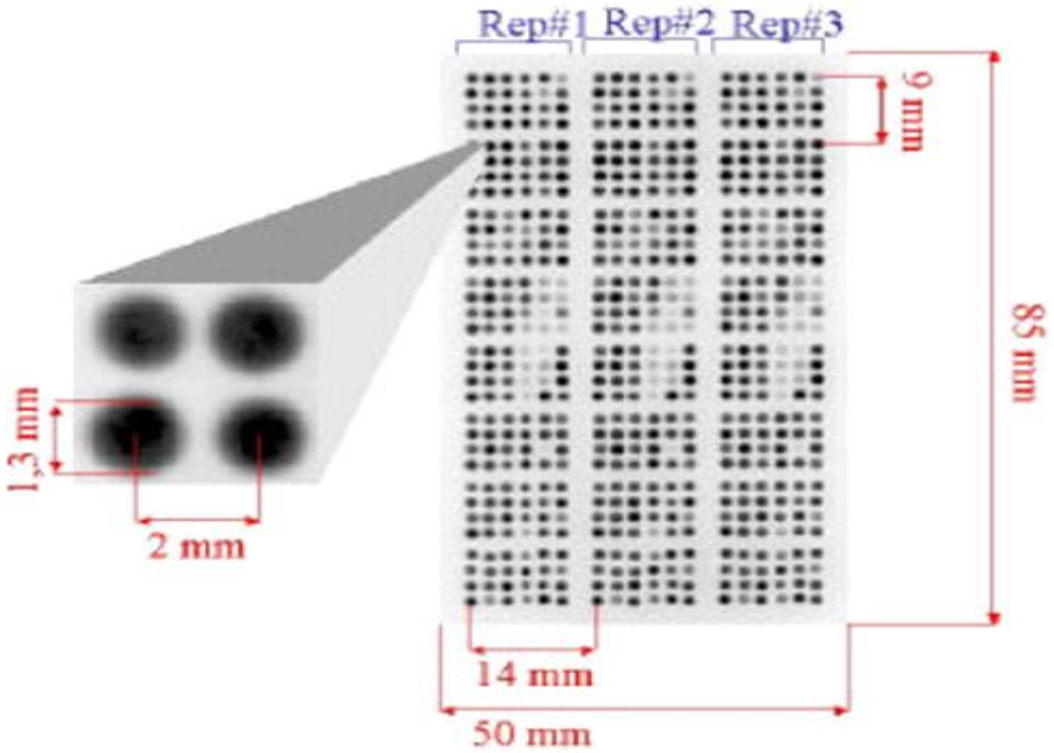


Fig-3: Microarray chip.

Table 1: Description of ArrayMining software as an example to understand Microarray data analysis

Software	Modules	Methods used by modules	Workflow
Array Mining.net	Cross-study normalisation	Linked gene and sample clustering method (XPN), an Empirical Bays method (EB), a median rank based score method (MRANK), an outliner- removing discritization technique (NorDi) and a Quantile discretization procedure (QDISC). While the first three methods provide continuous-valued outputs, the last two are based on discretization to filter out noise.	upload data in tab-delimited text file format or zip-compressed Affymetrix CEL file → normalized and summerized using the Robust Microarray Analysis Method → access to the GEO database on class discovery module → submit for task analysis → webpage containing the downloadable results generated
	Gene selection	The Empirical Bayes t-statistic (*eBayes*), the Significance Analysis in Microarrays method (SAM), a correlation-based combinatorial feature selection approach (CFS), a rankin method based on Random Forest classification (RF-MDA) and a Partial-Least-Squares based filter (PLS-CV)	
	Class discovery	The partition-based clustering methods available are *k-Means*, *PAM*, *SOM* and *SOTA*, and the hierarchical clustering methods are AverageLinkageAgglomerativeClustering, DivisiveAnalysis Clustering and Hybrid HierarchicalClustering	
	Class assignment	Various common methods for microarray sample classification	
	Gene set analysis	Our gene set analysis module provides access to three functional annotation sources to identify functionally related genes in a dataset and extract corresponding gene sets:The Gene Ontology Database, the KEGG data base, and a collection of 37 cancer-related gene sets from the van Andel Institute in Michigan	

7 million SAGE tags have been obtained. Despite these developments, only very few studies have employed this methodology for transcript profiling in higher plants, and the first reports on SAGE in the model plant species (*Arabidopsis*) appeared only very recently (Lee and Lee 2003). At present, a major limitation of SAGE is that in most species, tag to gene assignment (e.g. the identification of the gene the transcript of which has generated the SAGE tag) is based on EST clusters or on available cDNA sequences. This results in very incomplete identification of the transcripts revealed by SAGE tags, leaving many of them undetected in the databases.

SAGE protocol: SAGE is based mainly on two principles, representation mRNA (cDNAs) by short sequence tags and concentration of these tags to allow efficient sequence analysis. . The first principle is that a short oligonucleotide sequence, defined by a specific restriction endonuclease (anchoring enzyme, AE) at a fixed distance from the poly (A) tail, can uniquely identify mRNA transcripts. The second principle is that the end-to-end concatenation of these short oligonucleotides allows multiple transcript detection per sequencing reaction (Patino *et.al.,* 2002). The SAGE protocol starts with the purification of mRNA bound to solid phase oligo(dT) magnetic beads. The cDNA is synthesized directly on the oligo(dT) bead and then digested with the anchoring enzyme *Nla*III (AE) to reveal the 3'-most restriction site anchored to the oligo(dT) bead. Most SAGE experiments have used the 4-bp recognition site anchoring enzyme *Nla*III, predicted to occur every 256 bp and thus present on most mRNA species. However, creating a second SAGE library with a different anchoring enzyme may be useful for detecting transcripts without a *Nla*III site and also for reconfirming transcript identity in those with both anchoring restriction sites. This may significantly hamper data analysis, but the marginal utility of such an approach remains to be demonstrated. Next, the sample is equally divided into two separate tubes and ligated to two different linkers, A or B. Both linkers contain the recognition site for *Bsm*FI, a type IIS restriction enzyme that cuts 10-bp 3' from the anchoring enzyme recognition site. *Bsm*FI generates a unique oligonucleotide known as the SAGE tag, hence called the tagging enzyme (TE). The SAGE tags released from the oligo(dT) beads are then separated, blunted, and ligated to each other to give rise to ditags. The ditags are PCR amplified, released from the linkers, gel purified, serially ligated, cloned, and sequenced using an automated sequencer (Fig.4).

SAGE data analysis and followup strategies: The sequence files generated by the automated sequencer are analyzed using the SAGE2000 software (www.sagenet.org). The three steps involved in obtaining a differential gene expression list are as follows:

(1) Deciphering the SAGE tags from the sequence data files by using the SAGE2000 software for extracting ditags and checking for duplicate ditags;

Table 2: List of some software for microarray data analysis

Software Name	Access address	Remarks
Array Mining	http://www.arraymining.net/R-php-1/ASAP/microarrayinfobiotic.php	Online Microarray Data Mining
Cluster and Tree View	http://rana.lbl.gov/EisenSoftware.htm	Standard for hierarchical clustering and viewing dendrograms
GeneSpring GX	http://www.genomics.agilent.com/en/product.jsp?cid=AG-PT-130&tabId=AG-PR-1061&_requestid=2179534	Agilent's GeneSpring GX software provides powerful, accessible statistical tools for fast visualization and analysis of microarrays - expression arrays, miRNA, exon arrays and genomics copy number data
GeneCluster 2.0	http://www-genome.wi.mit.edu/cancer/software/genecluster2/gc2.html	Construct self-organizing maps, the latest version now also finds nearest neighbours
TM4	http://www.tm4.org	Microarray Data Manager (MADAM), TIGR_Spotfinder, Microarray Data Analysis System (MIDAS), and Multiexperiment Viewer (MeV), as well as a Minimal Information About a Microarray Experiment (MIAME)-compliantMySQL database, all of which are freely available to the scientific research community at TIGR's Software Download Site

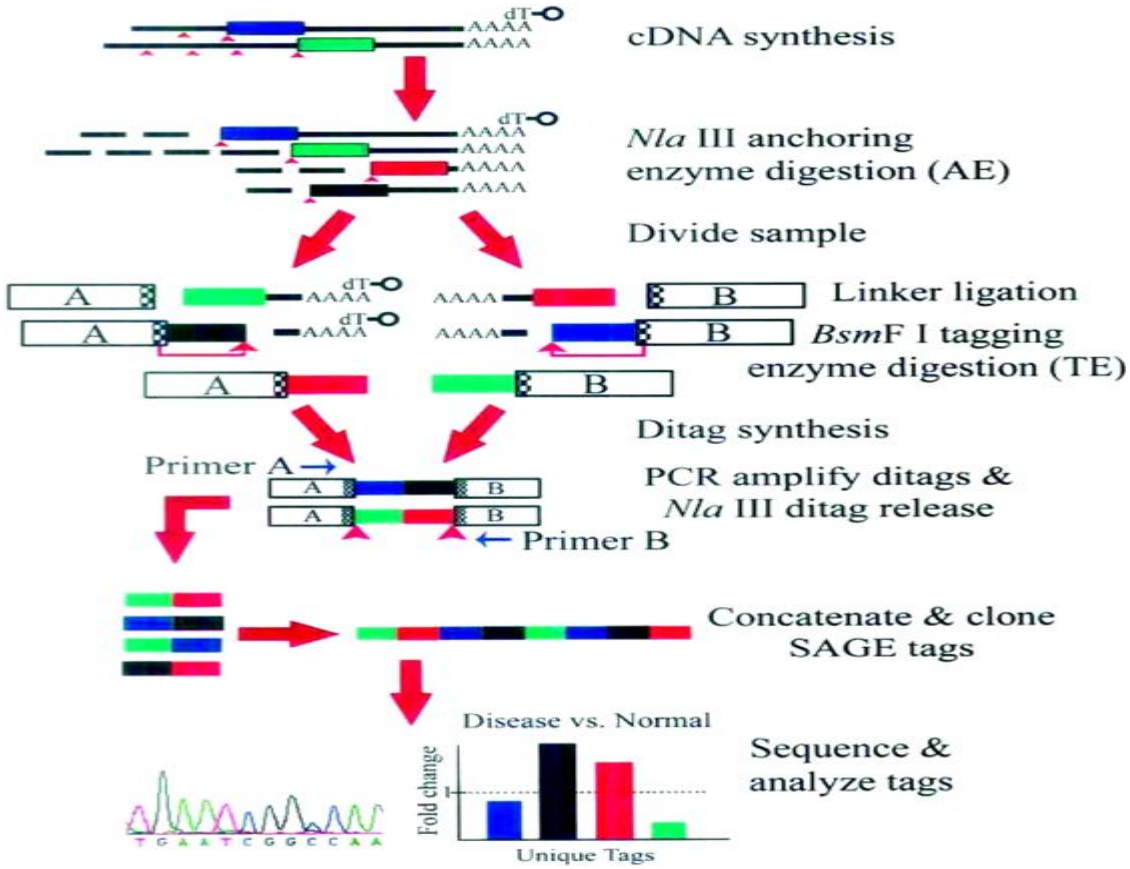


Fig-4: (Courtesy: Williom D. Patino, Omar Y. Mian and Paul M. Huang Serial Analysis of Gene Expression : Technical Considerations and Applications to Cardiovascular Biology, 2002 *cir.res* 91,565-569)

(2) Downloading a reference sequence database from the NCBI Web site (SAGEmap, www.ncbi.nlm.nih.gov);

(3) Associating the tags to the expressed gene database. The relative transcript abundance can then be calculated by dividing the unique tag count by the total tags sequenced, and the fold change can be determined by the ratio of tags between libraries. (Table3).

Massively Parallel Signature Sequencing (MPSS): MPSS is a recently developed high-throughput transcription proûling technology, has the ability to proûle almost every transcript in a sample without requiring prior Knowledge of the sequence of the transcribed genes. MPSS is one of the few technologies that produce data in a digital format. MPSS captured data by virtually counting all the mRNA in a tissue or cell sample. All genes are analysed simultaneously, and bioinformatics tools are used to study mRNAs (Brenner *et al*., 2000; Meyers *et al*., 2004).

Principle of MPSS analysis: Template sequences are determined by detecting successful adaptor ligations and a signature is obtained by monitoring a series of such ligations on the surface of a microbead in a fixed position in a flow cell. The sequencing method takes advantage of a special property of a type IIs restriction endonuclease; namely, its cleavage site is separated from its recognition site by a characteristic number of nucleotides (Bradford *et al*., 2010). Thus, a type IIs recognition site can be positioned in an adaptor so that after ligation, cleavage will occur inside the template to expose further bases for identification in the following cycle (Fig.5). Counting mRNA with MPSS is based on the ability to identify uniquely every mRNA in a sample. This is done by generating a 17-base sequence for each mRNA at a specific site upstream from its poly (A) tail (first DpnII site in double stranded cDNA). The 17-base sequence is then used as an mRNA identification signature. To measure the level of expression of any given gene, the total number of signatures for that gene mRNA

Cloning and sequencing cDNA fragments on beads: MPSS signatures for mRNAs in a sample are generated by sequencing dscDNA fragments cloned on microbeads. Complementary DNA (cDNA) is prepared from poly (A) RNA using a biotin labelled oligo- dT primer. The cDNA fragment is digested with DpnII (recognition sequence, GATC), and the 3'- most Dpn II poly A fragments are purified utilizing the biotin label at the end of each molecule. The fragments are subsequently cloned onto 5 micro meter diameter microbeads using a set 32 base tag/ anti tags. This process yields a library of beads where one starting mRNA molecule is represented by one microbead, and each microbead contains approximately 100,000 identical cDNA fragments from that mRNA. All molecules are covalently attached to the microbeads at their poly (A) ends, so the Dpn II end is available for the sequencing reactions.The sequencing process is initiated by ligation of an adapter molecule and digestion with a type II RE. Approximately one million microbeads are loaded into a specially designed flow cell in a way that allows them to stack together along channels and form a tightly packed monolayer in flow cell. The flow cell is connected to a computer controlled microfluidics network that delivers different reagents for the sequencing reaction. A high resolution CCD camera is positioned directly over the flow cell in order to capture

Table 3: List of some software for SAGE data analysis.

SOFTWARE	ACCESS ADDRESS	REMARKS
GermSAGE	http://germsage.nichd.nih.gov/	SAGE data on gene expression in male germ cell development.
5SAGE	http://5sage.gi.k.u-tokyo.ac.jp/	5'end serial analysis of gene expression.
SAGEmap	http://www.ncbi.nlm.nih.gov/SAGE	SAGEmap provides a tool for performing statistical tests designed specifically for differential-type analyses of SAGE (Serial Analysis of Gene Expression) data. The data include SAGE libraries generated by individual labs as well as those generated by the Cancer Genome Anatomy Project (CGAP), which have been submitted to Gene Expression Omnibus (GEO).
GOAL	http://microarrays.unife.it/	Gene Ontology Automated Lexicon (GOAL) is a tool for the functional analysis of data from SAGE and microarray experiments.
SAGExplore	http://protein.bio.puc.cl/cardex/servers/sagexplore/home.php	SAGExplore is a tool for the accurate mapping of experimental tags in serial analysis of gene expression (SAGE).
WebSage	http://bioserv.rpbs.jussieu.fr/websage/	WebSage is a tool that performs statistical analysis of SAGE data.

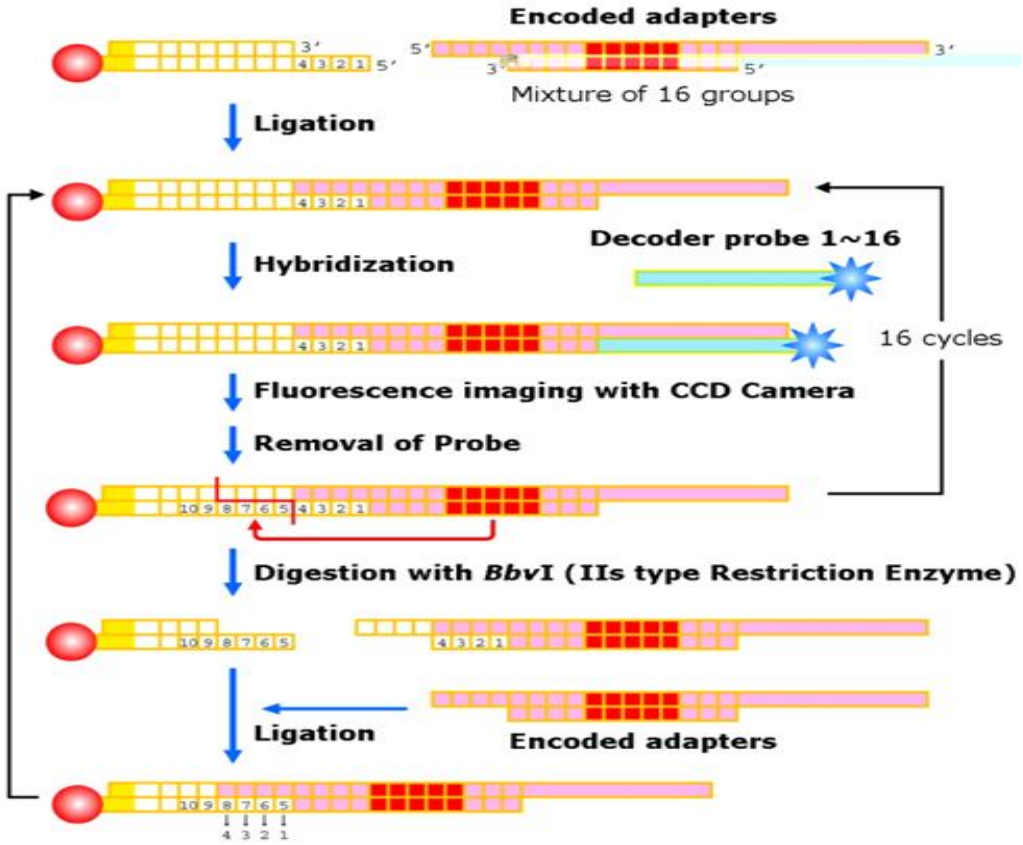


Fig-5: (Courtesy:BIOVIEW/www.takarabioeurope.com/custom service of comprehensive gene expression profiling through MPSS) Principle of MPSS sequencing

fluorescent images from the microbeads at specific stages of the sequencing reactions.

Data analysis

Each signature sequence in an MPSS dataset isanalyzed and compared with all other signatures. Identical signatures are counted. The level of expression of any single gene is calculated by dividing the number of signatures of all mRNA present in the dataset. The data for each gene is usually reported as the transcripts per million (TPM) (Cloonan *et. al*., 2008). Analysis of complete MPSS dataset makes it possible to calculate readily the genes that are expressed at varying levels within the sample (Table3).

RNA sequencing (RNA-Seq)

RNA sequencing or next generation sequencing (NGS) has emerged as a revolutionary tool in genetics, genomics, and epigenomics and holds promise in discovering *de novo* transcription/splice junctions and small RNAs with high specificity (Wu, 2013).While RNA-Seq is a relatively new method with high reproducibility and accuracy, it has already provided unprecedented insights into the transcriptional complexities of a variety of organisms, including yeast (Nagalakshmi *et. al*., 2008), mice (Mortazavi *et.al*., 2008), Arabidopsis (Eveland *et.al*., 2008) and humans(Sultan *et.al*., 2008).

Library preparation is a key step of RNA-seq, because it determines how closely the cDNA sequence data reûect the original RNA population. The most straightforward approach is to simply synthesise double-stranded cDNA, to which the adapter can be ligated (He *et. al*., 2008). To prepare high quality cDNAs, it is important to start with a population of intact mRNAs (Fig.2). Most eukaryotic mRNAs have several hundred bases of A at their 3' end. This poly A tail can be used to capture these mRNAs and remove contaminating rRNAs, tRNAs and other small cytoplasmic and nuclear RNAs. An oligo dT primer can be used with reverse transcriptase to make a DNA copy of the mRNA strand (Ingolia *et.al*., 2009). Alternatively, random primers can be used if one is searching for a particular mRNA or class of mRNAs. There are two general methods to convert RNA-DNA duplexes into cDNAs. In first approach the RNA strand is displaced or degraded, continue synthesis, after making a hairpin, until they have copied the entire DNA strand of the duplex. S1 nuclease can be used to cleave the hairpin and generate a cloning end. Unfortunately, the S1 nuclease treatment can also destroy some of the ends of the cDNA. An alternative procedure is to use RNase H to nick the RNA strands of the duplex. The resulting nicks can serve as primers for DNA polymerases like *E. coli* DNA

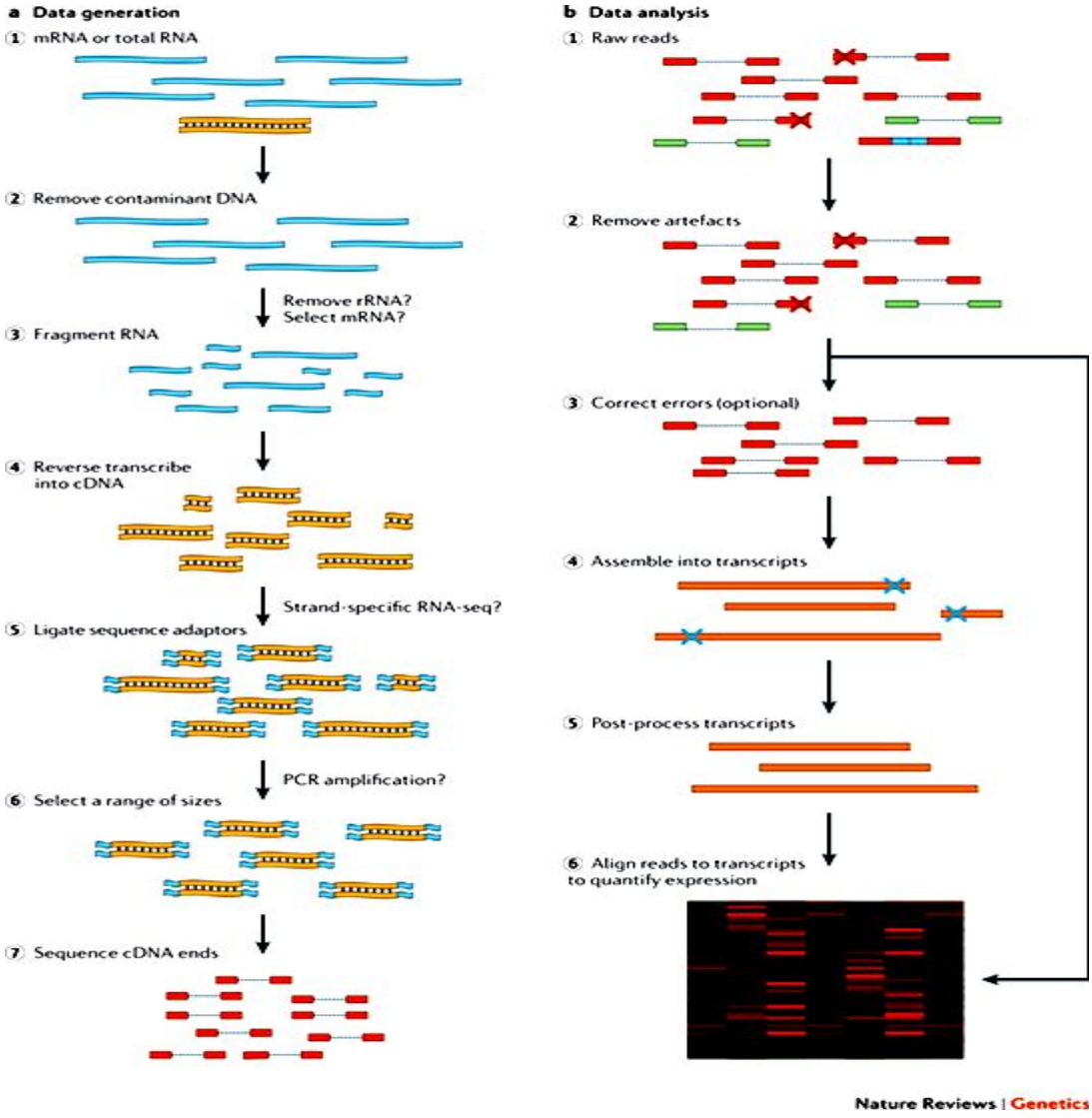


Fig-6: RNA-Seq and Data Analysis

polymerase I. This eventually leads to a complete DNA copy except for a few nicks which can be sealed by DNA ligases.

Two experimental protocols for RNA-Seq are in common use: (a) single end and (b) paired end sequencing experiments (Fig.6). For single end experiments, one end (typically about 50 to 100 bp) of a long (typically 200 to 400 nucleotide) molecule is sequenced. For paired end experiments, typically 50–100 bp of both ends of a typically 200 to 400 nucleotide molecule are sequenced (Wang *et. al*., 2009). Using current Illumina technology, each time the sequencing machine is operated, eight samples (e.g., potentially eight diûerent catalogues of gene expression) can be interrogated (essentially) independently and tens of millions of reads are produced in each sample.

RNA-Seq data analysis: Once high-quality reads have been obtained, the first task of data analysis is to map the short reads from RNA-Seq to the reference genome, or to assemble them into contigs before aligning them to the genomic sequence to reveal transcription (Fig.4) structure (Jiang and Wong, 2009, Mortazavi *et al*., 2008). There are several programs for mapping reads to the genome, including ELAND, SOAP31, MAQ32 and RMAP. However, short transcriptomic reads also contain reads that span exon junctions or that contains poly (A) ends - these cannot be analysed in the same way. For genomes in which splicing is rare (for example, *S. cerevisiae*) special attention only needs to be given to poly (A) tails and to a small number of exon–exon junctions. Poly (A) tails can be identified simply by the presence of multiple As or Ts at the end of some reads. Exon–exon junctions can be identified by the presence of a specific sequence context (GT–AG dinucleotides that flank splice sites) and confirmed by the low expression of intronic

Table 4: List of some open source solution for RNA-Seq Data analysis

Software Name	Access address	Remarks
Array Mining	http://www.arraymining.net/R-php-1/ASAP/microarrayinfobiotic.php	Online Microarray Data Mining
Cluster and Tree View	http://rana.lbl.gov/EisenSoftware.htm	Standard for hierarchical clustering and viewing dendrograms
Gene Spring GX	http://www.genomics.agilent.com/en/product.jsp?cid=AG-PT-130&tabId=AG-PR-1061&_requestid=2179534	Agilent's GeneSpring GX software provides powerful, accessible statistical tools for fast visualization and analysis of microarrays - expression arrays, miRNA, exon arrays and genomics copy number data
Gene Cluster 2.0	http://www-genome.wi.mit.edu/cancer/software/genecluster2/gc2.html	Construct self-organizing maps, the latest version now also finds nearest neighbours
TM4	http://www.tm4.org	Microarray Data Manager (MADAM), TIGR_Spotfinder, Microarray Data Analysis System (MIDAS), and Multiexperiment Viewer (MeV), as well as a Minimal Information About a Microarray Experiment (MIAME compliant MySQL database, all of which are freely available to the scientific research community at TIGR's Software Download Site

sequences, which are removed during splicing. Transcriptome maps have been generated in this manner for *S. cerevisiae* (Wang *et al*., 2009). For complex transcriptomes it is more difficult to map reads that span splice junctions, owing to the presence of extensive AS and *trans*-splicing. One partial solution is to compile a junction library that contains all the found junction sequences and map reads to this library. A challenge for the future is to develop computationally simple methods to identify novel splicing events that take place between two distant sequences or between exons from two different genes. (Table 4).

Application of transcriptome sequencing to marker discovery in plants

Genetic variation within commercialized crop varieties is not usually well characterized or quantified. It follows then that the effect of intra-varietal genetic variation on crop performance under stress is also poorly understood, which may put production at risk from changing climate and rapidly evolving pests and diseases. Transcriptome sequencing allows genome-wide analysis of large, complex plant genomes and the potential to identify biologically significant SNPs. The genetic variation between and within barley varieties was defined by deep sequencing and assembled into unigenes the transcriptomes of two barley varieties Baudin and Gairdner (Henry *et. al*., 2012). A large number of SNPs were identified, with more than 200,000 SNP between DNA sequence reads for variety Baudin and reference EST sequences, and more than 300,000 SNP between Baudin reads and reads from the variety Gairdner. Significant SNPs (SNP allele frequency > 0.1) represented 9.65% for Baudin and 14.64% for Gairdner genetic variation. Background genetic diversity (SNP allele frequency d" 0.1) accounted for 90.23% and 85.52% of genetic variation in Baudin and Gairdner, respectively. The SNP dataset was further refined to produce a set of very high-quality SNPs for varietal genotyping. Although SNP variation within varieties has not been widely examined in other species, analyses of SNPs between varieties have been undertaken to facilitate varietal distinction in many plant species like wheat, rice (Gopala Krishnan *et. al*., 2012), maize (Barbazuk *et. al*., 2007), chickpea (Hiremath *et. al*., 2011), pigeonpea (Dubey *et. al*., 2011), soybean (Wu *et. al*., 2010) and oilseed rape (Trick *et. al*., 2009). These proves that markers developed by transcriptome sequencing technologies provide an unprecedented understanding of the levels of genetic variation in plants which become a valuable tool for plant breeders for unique selection of diversity within varieties.

CONCLUSION

All the methods discussed above are high-throughput to profile the transcriptome. Sequencing based techniques (RNA-seq, MPSS and SAGE) can provide complete transcriptional characterization of all the cells of an organism while hybridization based techniques produce much significant information about deployed transcriptome in different cell types and tissues, how gene expression changes across development states and how it varies within and between species. Sequencing transcripts (*that is,* expressed genes) is inherently cheaper than sequencing genomes, because it eliminates the need to sequence the intronic and intergenic regions, which can be orders of magnitude larger. From this information one can generate new hypotheses about biology or test existing ones. The size

and complexity of these experiments often results in a wide variety of possible interpretations. Good experimental design, adequate biological replication and follow up experiments play key roles in successful expression profiling experiments.

REFERENCES

Barbazuk, W.B., Emrich, S.J., Chen, L.L., Schnable, P.S. (2007). SNP discovery via 454 transcriptome sequencing. *Plant Journal,* **51**: 910–918.

Berget, S.M., Moore, C., Sharp, P.A. (1977).Spliced segments at the 52 terminus of adenovirus 2 late mRNA. *Proceedings of Natural Acadamic Science,* **74**:3171–3175.

Bradford, J.R., Hey, Y., Yates, T., Li, Y., Pepper, S.D., Miller, C.J. (2010). A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics,* **11**: 282-294.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., *et al.* (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*. **18**: 630-634.

Brett, D., Pospisil, H., Valcarcel, J., Reich, J., Bork, P. (2002).Alternative splicing and genome complexity. *Nature Genetics,* **30**:29–30.

Byers, R.J., Hoyland, J.A., Dixon, J., Freemont, A.J. (2002). Subtractive hybridization -genetic takeaways and the search for meaning. *International Journal of Experimental Pathology,* **81**: 391-404.

Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods,* 5 (**7**): 613 – 619.

Danila, A.L., Laborde, L., Legrand, S., Huot, L., Hot, D., Lemoine, Y., Hilbert, J.L., *et al.* (2010). (Identification of novel genes potentially involved in somatic embryogenesis in chicory (*Cichorium intybus* L.). *BMC Plant Biology,* **10**: 122-137.

Dubey, A., Farmer, A., Schlueter, J., Cannon, S.B., Abernathy, B., Tuteja, R., Woodward, J., Shah, T., *et al*. (2011). Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). *DNA Research,* **18**: 153–164.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell,* **20**:313–319.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceeding of Natural Acadamy of Science,* **95**: 14863–14868.

Eveland AL, McCarty DR, Koch KE (2008) Transcript profiling by 32 -untranslated region sequencing resolves expression of gene families. *Plant Physiol.* 146:32–44.

Gopalakrishnan S, Upadhyaya HD, Vadlamudi S, Humayun P, Vidya MS, Alekhya G, *et al.* (2012) Plant growth-promoting traits of biocontrol potential bacteria isolated from rice rhizosphere. Springerplus 1:71.

Harrington, C.A., Rosenow, C., Retief, J. (2000).Monitoring gene expression using DNA microarrays.*Current Opinion in Microbiology,* **3**:285–291.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., Kinzler, K.W. (2008). The antisense transcriptomes of human cells. *Science,* **322**:1855–1857.

Henry RJ, Edwards M, Waters DLE, GopalaKrishnan S, Bundock P, Sexton TR, Masouleh AK, Nock CJ, Pattemore J (2012) Application of large-scale sequencing to marker discovery in plants. *Biosciences J.* **37**(5): 829-841.

Hiremath, P.J., Farmer, A., Cannon, S.B., Woodward, J., Kudapa, H., Tuteja, R., Kumar, A., BhanuPrakash, A., *et al*. (2011). Large-scale transcriptome analysis of chickpea (*Cicer arietinum* L.) an orphan legume crop of the semi-arid tropics of Asia and Africa. *Journal of Plant Biotechnology,* **9**:922–931.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome *profiling. Science* **324**:218–223.

Jiang, H., and Wong, W.H. (2009). Statistical inferences for isoform expression in RNA-Seq. *Bioinfo*. **25**(8): 1026-1032

Jiang, Y., Harlocker, S.L., Molesh, D.A., Dillon, D.C., Houghton, R.L., Repasky, E.A. *et al.* (2002). Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays. *Oncogene*, **21**:2270 – 2282.

Kim, E., Magen, A., Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Reearch,* **35**:125–131.

Lee, J.Y., Lee, D.H. (2003). Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiology*, **132**: 517-529.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods* 7(**9**): 709–715.

Lievens S, Goormachtig S, Holsters M (2001) A critical evaluation of differential display as a tool to identify genes involved in legume nodulation: looking back and looking forward. *Nucleic Acids Res* **17**: 3459–3468.

Meyers, B.C., Lee, D.K., Vu, T.H., Tej, S.S., Edberg, S.B., Matvienko, M. ,Tindell, L.D. (2004). Arabidopsis MPSS: An online resource for quantitative expression analysis. *Plant Physiology,* **135**: 801–813.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**:621–628.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**(5881):1344-1349.

Patino, W.D., Mian, O.Y., Hwang, P.M. (2002). Serial analysis of gene expression technical considerations and applications to cardiovascular biology. *Circular research,* **91**: 565-569.

Reddy, A.S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol*. **58**:267–294.

Rosenfeld, M.G., Lin, C.R., Amara, S.G., Stolarsky, L., Roos, B.A., Ong, E.S., Evans, R.M. (1982). Calcitonin mRNA polymorphism: Peptide switching associated with alternative RNA splicing events. *Proceedings of Natural and Academic Science,* **79**:1717–1721.

Sharp, P.A. (1994). Split genes and RNA splicing. *Cell,* **77:** 805–815.

Sorek, R., Ast, G. (2003). Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Research,* **13**:1631–1637.

Staley,J.P., Guthrie, C. (1998). Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell*, **92**:315–326.

Sultan, M., Schulz, M.H., Richard, H., *et. al.* (2008). A Global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science,* 321(**5891**): 956-960.

Trick, M., Long, Y., Meng, J., Bancroft, I. (2009). Single nucleotide polymorphism (SNP) discovery in the polyploidy Brassica napus using Solexa transcriptome sequencing. *Journal of Plant Biotechnology,* **7**:334–346.

Virlon, B., Cheval, L., Buhler, J.M., Billon, E., Doucet, A.J., Elalouf, J.M. (1999). Serial microanalysis of renal transcriptomes. *Proceedings of Natural and Academic Science,* **96**:5286–15291.

Wang, B.B. and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *PNAS*. 103(18):7175-7180.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Review Genetics,* 10(**1**):57–63.

Wu, M., Tu, T., Huang, Y., Wu, Y.C. (2013). Suppression subtractive hybridization identified differentially expressed genes in lung adenocarcinoma: ERGIC3 as a novel lung cancerrelated gene. *BMC Cancer,* **13**:44-54.

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

Xing, Y. and Lee, C. (2006). Alternative splicing and RNA selection pressure - evolutionary consequences for eukaryotic genomes. *Nature Review Genetics*, **7**:499–509.