MASSIVELY PARALLEL SIGNATURE SEQUENCING

Dr. Mrinmoy Choudhury

M.V.Sc. Scholar

Department of Animal Biotechnology
College of Veterinary Science, AAU, Khanapara, Guwahati- 781022

Email id: mrinmoych02@gmail.com

Introduction

Genomic studies are efficiently done nowadays due to the availability of recent technologies of sequencing a genome or gene in a faster and an accurate way. If there are datasets containing gene expression profiles, it would provide a big contribution in experiments concerning digital biology. One such tool was designed for the mentioned purpose called Massively Parallel Signature Sequencing or MPSS. MPSS is a new tool for in-depth profiling of quantitative gene expression in all organisms. It analyses the expression level of all genes in a sample by counting the number of mRNA molecules synthesized from each gene. The genes are analyzed and with the help of bioinformatics tools, the mRNA produced from each gene are sorted out relative to the total number of molecules in the sample. MPSS forms a dataset which simplifies the management and analysis of the data. It can be utilized to generate complete datasets which will help in facilitating hypothesis- driven experiments.

How do MPSS works

MPSS has the ability to uniquely identify all the mRNA in the sample. A 17- base sequence for each mRNA is first generated at a specific site upstream from its poly-A tail. The 17- base sequence will be used as a 'signature' to identify the mRNA in the sample. Hence, when gene

expression level is measured, MPSS will count the total number of signatures for that particular gene's mRNA. With the help of biotin - labelled oligo-dT primer, complementary DNA (cDNA) is prepared from poly-A tail. The oligo-dT is designed to prime each mRNA molecule exactly at poly–A tail. The cDNA fragments are digested with *Dpn*II restriction endonuclease enzyme (recognition sequence GATC), and the poly-A tail fragments are purified using biotin label at the end of each molecule. The cDNA molecules are then cloned into a plasmid vector containing 32 base pair oligonucleotide tags. The clones are then amplified by polymerase chain reaction (PCR), and these amplicons are treated with exonucleases to form single- strands of 32 – base tags, which will be present at the end of each cDNA molecules. These tags are hybridized to 32 – base complementary tags which are covalently linked to 5 µm microbeads. For every 32 – base tag in cDNA, there will be 32 – base complementary tag in the microbeads. After hybridization, the nicks between cDNA and the microbeads are ligated with the help of enzymes. This process will lead to formation of a microbead library, where one mRNA is represented by one microbead and each microbead contains 100,000 cDNA fragments which are identical to the cDNA produced from the mRNA.

The next step is the sequencing in which the single- stranded overhangs produced by *Dpn*II are ligated to adaptor molecule. The sequences are then digested using *Bbv*I, which nicks the DNA 9 – 13 positions away from the recognition sequence. This leads to production of molecules with four – base single - stranded overhangs, which are adjacent to *Dpn*II sequence. Another adapter called encoded adaptor is also used during sequencing, which contains four - base single – stranded overhangs at one end and these overhangs contain all possible nucleotide combinations. These overhangs are hybridized to the cDNA microbeads, and it is detected through 16 hybridization reactions with the help of fluorescent decoder probes. This process will yield first four nucleotides at the end of each molecule. *Bbv*I will digest the encoded adaptor from the first round, and the whole process will be repeated to generate a 17 – base signature sequence.

The gene expression level or RNA abundance is calculated after the sequencing process. Each of the signature sequences in the MPSS dataset is analyzed and compared with the identical and non – identical signatures in the dataset. The level of gene expression is measured by

dividing the number of signatures of a gene by the total number of signatures of all the mRNAs present in the dataset. The data is measured as transcripts per million (TPM).

Advantages of MPSS

1. MPSS can sequence one million cDNA clones simultaneously.

2. MPSS is complementary with microarrays for biological experiments.

3. MPSS can map higher percentages of signature sequences to unique locations as compared to SAGE.

4. The automated nature of MPSS can produce large number of datasets of signature sequences.

5. MPSS can help to differentiate homologous genes from each other because the signature sequence remains in the 3´ untranslated region, which is different for each individual gene.

6. MPSS enables in – depth profiling of gene expression in all organisms.

Conclusion

Several other technologies are being used daily for genome or gene sequencing. These technologies may be analogous, where the level of gene expression is measured through the process of hybridization of labeled probes to a solid support. Other than MPSS, SAGE or cDNA technologies are preferred for measuring gene expression level, but they have certain limitations. MPSS datasets are proficient for building relational databases for biological experiments. It can be used for statistical analysis by capitalizing large number of measurements in the signature datasets, and helps to know whether a gene can be expressed differently between two samples. In MPSS, there is high chance of reproducibility of data. As compared to other gene sequencing technologies, it can be cheaper and faster, and it can simultaneously clone millions of copies of cDNA from a particular gene sequence. MPSS does not require any prior knowledge of gene or genome and as a result, it is possible to produce quantitative datasets of gene expression from all organisms. It is sensitive and amongst the million copies of cDNA, it can accurately measure the gene expression level of genes that are expressed at low level in the cell.

References

Brenner, S.; Johnson, M.; Bridgham, J.; Golda, G.; Lloyd, D.H.; Johnson, D.; Luo, S.; McCurdy, S.; Foy, M.; Ewan, M. and Roth, R. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature biotechnology*. **18**(6): 630-634.

Lockhart, D.J.; Dong, H.; Byrne, M.C.; Follettie, M.T.; Gallo, M.V.; Chee, M.S.; Mittmann, M.; Wang, C.; Kobayashi, M.; Norton, H. and Brown, E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature biotechnology*. **14**(13): 1675-1680.

Reinartz, J.; Bruyns, E.; Lin, J.Z.; Burcham, T.; Brenner, S.; Bowen, B.; Kramer, M. and Woychik, R. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics*. **1**(1): 95-104.

Tyagi, S. (2000). Taking a census of mRNA populations with microbeads. *Nature biotechnology*. **18**(6): 597-598.