International Journal of Engineering Management and Life Sciences (IJEMLS)
Volume: 01 Issue: 05 | Octoer-2016 **ISSN: 2456-2580** **www.ijemls.com**

Expressed Sequence Tags: Making Gene Discovery Easier

Chinedu Ugwoke, Okonkwo Umeude, Emmanuel Asogwa

College of Natural and Applied Sciences, Crescent University, Abeokuta, Ogun State, Nigeria

Abstract

Expressed Sequence Tags (ESTs) are expressed segments of individual's genome. ESTs represent only the fragments of genes, not complete coding sequence. Some ESTs projects are made by concentrating on 5' end in order to maximize the amount of coding sequence determined. But since EST sequence is generated only once, the sequence may contain errors. ESTs are identified by their clone number and their 5' to 3' orientation in the databases. This article explores various functional aspects of ESTs.

Key Words: ESTs, Expression, DNA, Discovery, Gene

INTRODUCTION

Researchers are thriving diligently to sequence and assemble the genomes of different organisms, including the mice and man, for a number of vital reasons. Although important aims of any sequencing project is to obtain a genomic sequence and find out a complete set of genes, the ultimate goal is to acquire an understanding of where, when and how a gene is turned on and off, a process commonly known as gene expression. Once we begin to understand when, where and how a gene is expressed under normal conditions, we can then study what happens in an altered state, such as in illness. To accomplish the latter goal, however, investigators must identify and study the protein, or proteins, coded for by a gene [1].

As one can imagine, finding a gene that codes for a protein, or proteins, is not easy. Traditionally, scientists would start their search by defining a biological problem and developing a strategy for researching the problem. Oftentimes, a search of the scientific literature provided various clues about how to proceed. For example, other laboratories may have published data that established a link between a particular protein and a disease of interest. Researchers would then work to isolate that protein, determine its function, and locate the gene that coded for the protein. Alternatively, scientists could conduct what is referred to as linkage studies to determine the chromosomal location of a particular gene. Once the chromosomal location was determined, scientists would use biochemical methods to isolate the gene and its corresponding protein. Either way, these methods took a great deal of time—years in some cases—and yielded the location and description of only a small percentage of the genes found in the human genome.

Now, however, the time required to locate and fully describe a gene is rapidly decreasing, thanks to the development of, and access to, a technology used to generate what are called Expressed Sequence Tags, or ESTs. ESTs provide investigators with a fast and inexpensive route for discovering new genes, for obtaining data on gene expression and regulation, and for generating genome maps. Today, researchers using ESTs to study the human genome find themselves riding the crest of a wave of scientific discovery the likes of which was not seen earlier.

ESTs are tiny pieces of DNA sequence (usually 150 to 600 nucleotides lengthwise) that are produced by sequencing at one end or both ends of an expressed gene [2]. The logic is to sequence bits of DNA that represent genes expressed in certain cells, tissues, or organs from different organisms and use these 'tags' to find a gene out of a portion of chromosomal DNA by finding similar base pairs. The difficulty associated with identifying genes from genomic sequences varies among organisms and is dependent upon genome size as well as the presence or absence of introns, the intervening DNA sequences interrupting the protein coding sequence of a gene.

Gene identification is very difficult in humans, because most of our genome is composed of introns interspersed with a relative few DNA coding sequences, or genes. These genes are expressed as proteins, a complex process composed of two main two steps. Each gene (DNA) must be converted, or transcribed, into messenger RNA (mRNA), RNA that serves as a template for protein synthesis. The resulting mRNA then guides the synthesis of a protein through a process called translation. Interestingly, mRNAs in a cell do not contain sequences from the regions between genes, nor from the non-coding introns that are

present within many genes. Therefore, isolating mRNA is key to finding expressed genes in the vast expanse of the human genome. The problem, however, is that mRNA is very unstable outside of a cell; therefore, scientists use special enzymes to convert it to complementary DNA (cDNA). cDNA is a much more stable compound and, importantly, because it was generated from a mRNA in which the introns have been removed, cDNA represents only expressed DNA sequence [3].

GENERATION OF ESTs THROUGH cDNA

Once cDNA representing an expressed gene has been isolated, scientists can then sequence a few hundred nucleotides from either end of the molecule to create two different kinds of ESTs. Sequencing only the beginning portion of the cDNA produces what is called a 5' EST. A 5' EST is obtained from the portion of a transcript that usually codes for a protein [4]. These regions tend to be conserved across species and do not change much within a gene family. Sequencing the ending portion of the cDNA molecule produces what is called a 3' EST. Because these ESTs are generated from the 3' end of a transcript, they are likely to fall within non-coding, or untranslated regions (UTRs), and therefore tend to exhibit less cross-species conservation than do coding sequences.

APPLICATIONS OF ESTs

Just as a person driving a car may need a map to find a destination, scientists searching for genes also need genome maps to help them to navigate through the billions of nucleotides that make up the human genome. For a map to make navigational sense, it must include reliable landmarks or "markers". Currently, the most powerful mapping technique, and one that has been used to generate many genome maps, relies on Sequence Tagged Site (STS) mapping. An STS is a short DNA sequence that is easily recognizable and occurs only once in a genome (or chromosome). The 3' ESTs serve as a common source of STSs because of their likelihood of being unique to a particular species and provide the additional feature of pointing directly to an expressed gene.

Because ESTs represent a copy of just the interesting part of a genome, that which is expressed, they have proven themselves again and again as powerful tools in the hunt for genes involved in hereditary diseases. ESTs also have a number of practical advantages in that their sequences can be generated rapidly and inexpensively, only one sequencing experiment is needed per each cDNA generated, and they do not have to be checked for sequencing errors because mistakes do not prevent identification of the gene from which the EST was derived [5].

To find a disease gene using this approach, scientists first use observable biological clues to identify ESTs that may correspond to disease gene candidates. Scientists then examine the DNA of disease patients for mutations in one or more of these candidate genes to confirm gene identity. Using this method, scientists have already isolated genes involved in Alzheimer's disease, colon cancer, and many other diseases. It is easy to see why ESTs will pave the way to new horizons in genetic research.

CONCLUSION

The EST sequences are source of gene-targeted and tissue-specific markers. They facilitate the fabrication of DNA array for genetic studies. They serve as markers for various diseases. ESTs may potentially serve as an important tool for comparative genomic studies.

CONFLICT OF INTEREST: None

REFERENCES

[1] Chou H, Holmes MH (2001). DNA sequence quality trimming and vector removal. Bioinformatics. 17: 1093-1104

[2] Huang X, Madan A: CAP3 (1999). A DNA sequence assembly program. Genome Res. 9: 868-77

[3] Stekel DJ, Git Y, Falciani F (2000). The comparison of gene expression from multiple cDNA libraries. Genome Res. 10: 2055-61.

[4] Flood J, Keenan L, Wayne S, Hasan Y (2005). Studies on oil palm trunks as sources of infection in the field. Mycopathologia. 159: 101-07

[5] Romualdi C, Bortoluzzi S, d'Alessi F et sl. (2003). IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. Physiol Genomics. 12: 159-62

ISSN 2456-2580