Functional Genomics Towards Understanding of the Mechanisms of Life

Yuji Kohara

National Institute of Genetics

My name is Yuji Kohara and I am with the National Institute of Genetics. We are participating in some joint research of so called "environmental hormones," but this is not our specialty, so we were somewhat thrown off by it. Be that as it may, today I would like to talk about the present situation of the function genomics of *C. elegans*.

Of all multicellular organisms, the genomes of *C. elegans* are determined most quickly. Such information concerning genomes has been shared among the scientific community for some time now, and it is because of this that genome research has progressed. The fruit fly has subsequently been determined. Roughly speaking, human beings have also been determined. As I am sure you are aware, number of genes has become a serious problem recently. We were very surprised to find for instance that flies have fewer genes than do nematodes. Genomes are however ingenuously constructed, performing some function depending on how the various genes are worked. We want to know the mechanism by which this works.

Looking back upon evolution, we think the structure of genomes was like the prototype of a multicellular organism that mutated. The entire genome was duplicated possibly two times, followed by further mutation, eventually evolving to its current status. Just as Dr. Inoue said a little while ago, I also think that comparing various species of genomes is the fastest way to understand organisms. This is why genome research is progressing. In this respect, we could say that flies and nematodes are the prototypes of human genomes. Like tubifex, *C. elegans* subsists on bacteria. The approximately 1 millimeter-long organism propels itself in a snaking motion. There are many types of nematodes and we are using *C. elegans* among them. The organism's body consists of approximately 1000 cells, of which the way they divide is strictly fixed. Approximately one-third of those cells is nerve cells. Although extremely simple, as an animal, the basic pattern is triblastica. It therefore serves as a model animal.

The organism's body is transparent, so the entire process can be observed under a microscope as you see. Here we see fertilization taking place, the eggs dividing, and gradually the cell divides. During this time, the cells divide and the intestines are formed in this area. The muscles decompose in this area and the mouth is formed. That is how the process occurs. Here we have speeded up the events of about eight hours.

With modern technologies using GFP – here is a Histon GFP – we can observe living genome movement and distribution. By effecting mutation here, we can see the process by which various cells divide and subtle effects.

Unlike human beings, with the first division – in the case of people, twins result if both divide – in the case of nematodes, the fate of the front and rear is completely different. The point is in what way this will change. In order to get the full picture, after the genome structure is determined, we establish the genetic structure and systematically clarify when and where each gene will be generated, what kind of cell will it become and be used and what it will do. This is the main point of function genomics. The ultimate target is to reproduce generation by computer, and that is why we are attempting to obtain data.

Thus in order to gradually obtain genes at first, we are randomly analyzing cDNA as part of the so-called "EST Project." We have already identified cDNA of about 10,000 of the approximately 19,000

predicted genes. Using this, Exon/Intron boundary is determined, genetic structure is established and prediction is gradually corrected by comparing with genomes.

Here is a database called "WormBase" prepared by Caltech. The database contains all genome data and phenotype data.

Concerning manifestation, we are attempting to clarify in detail when and by what cells is synthesis of messenger RNA (mRNA), which is the initial event of *genes*, caused. Nematodes are extremely small and thin, so we are systematically attempting to learn about mRNA by applying the *in situ* hybridization using the entire organism. There are 10,000 applicable genes, and common sense tells us it would impractical to do this for all of them. Instead we have prepared a simple method and are proceeding as follows.

Here is a typical pattern. We can see various stages of the embryo in a single view. In the case of this gene, we can see messengers from two cells. This is maternal manifestation; maternal manifestation ceases as generation progresses. It exhibits the pattern. This sort of phenomena frequently occurs. Because there are 10,000 genes, we are training technicians, gradually obtaining data and are promoting the use of annotation.

This part is extremely difficult. We have arranged embryogenesis in time sequence from this for instance.

Here we have the result of doing this for five different genes. Genes for example are as such, and in fact, are specifically manifested in the hypodermis. For the second stage, the genes are specifically manifested in muscle cells (body of muscle) to an extreme degree. These three on the other hand are manifested by genealogical cells, all of which form intestines. These genes are not present at the beginning, but first appear during this stage. We could therefore say that transcription begins in this stage. This occurs much later for gene at the bottom. This is produced by cells dividing, and occurs at the end of the time sequence.

The act of arranging such patterns for about 7500 genes as of the present – we have in fact arranged 7500 lines with annotation – as you can see, lots of manifestation points are given. We are executing so-called "gene manifestation clustering" and gathering groups of similar manifestation pattern one after another.

Thus, we gather specifically manifested genes, such as this cell lineage by which intestines are formed. This is what we get when this is sorted in the order of manifestation. Transcription begins in an extremely early stage in this group of genes. We can now define the time of the start of transcription, which we couldn't do before.

With this group of genes on the other hand, it appears around here not in the time of the start. It however appears in this offspring cell of the same cell. Oppositely each group of genes are regulated differently by the various individual cells.

If we map this, for example this stage, this gene indicated in light blue is in this area of the genome. Transcription really starts during this stage. Transcription starts for the gene indicated in pink during this stage. This is the stage called "gastrulation." This sort first starts during this stage. Similarly, genes indicated in green are manifested in sequence for the same offspring cell and ignite in sequence as you see here. In this way, we see that tissue – intestines – in this case – is formed.

The problem is what way each gene is regulated. Here they may be regulated the same way. We therefore search for the motif for being regulated. It is of course important to do this experimentally, but research is progressing in terms of information science.

This is a specific gene appearing for genealogy of intestines. With this as a positive group – although it appears in other places – it does not appear in this red area. Groups of genes such as this are gathered as the negative group. In the upper class of these genes, but not here – we searched for such a

motif in terms of information science. Motifs are extremely short, so we will study various sizes such as 7 base, 8 base and 9 base. This is called “window search.” We will continue to trace this. Zero – in this case, the starting line of transcription is more or less arranged.

There is much for positive, and for negative, we search for none if possible. If the motif is short, however, naturally it appears as a fixed probability. These three sequence motifs always appeared for positive groups. Although it’s a little hard to see, it appears to a certain extent for the lower negative group. Having these three at the same time under these conditions -- they are “linked” in other words – if we do this, all disappears below with only the positive above remaining. This is therefore very significant in terms of statistics. It is highly possible that the combination of these three motifs is extremely important for gene manifestation. Although this needs to be checked out experimentally, this sort of data has been extracted.

This is another motif search of by cells that form muscle, and as you might expect, this is what we find. Naturally, if this is done individually, it will result in negative data. However, when this is linked for sharing, all will disappear for negative, and only positive will remain. If we search the database for this motif -- for example this pink one – this sometimes strongly resembles the binding site of the factor that determines division of muscle cells, so we could say that the results are extremely consistent. Doing this, we are now systematically searching the regulatory motif of the gene groups showing specific genealogical cell manifestation.

Concerning the manifestation pattern, as you might expect we are using micro array. Although it has been around for some time, with micro array, about 10,000 cDNA are contained on a membrane, and we compare using messengers of different conditions.

This is joint research with Jonathan Eubank of France. We can infect a nematode with a certain type of bacteria. He is monitoring the specimen to see what happens. These are the results of a study to see how the gene manifestation changes when the specimen is infected, a short time after being infected and some time after being infected. We found that the genes increase.

We already know all of the gene manifestation patterns, so when this appears, we know right away what it is. These are cells of the root of the mouth and intestines; they are strongly manifesting here. We also found that they are manifested in the intestines. Thus we can link the manifestation array and function inference. Of course DNA chips and micro arrays have been prepared and can be used.

Last of all is function. In the case of nematodes, there is the RNAi method shown here. When we inject two-chain RNA corresponding to exon, the corresponding messenger is destroyed, enabling gene knockout. In the case of nematodes, you can impregnate the parent insect or just immerse the insect in an RNA solution. Two to three days after impregnation, by observing the phenotype of the resultant offspring, you can find out the knockout phenotype, so it is an excellent method.

Up to now – this is joint research with Asako Sugimoto of the University of Tokyo – of the series of random cDNA, these are the results of RNAi from one end concerning approximately 2600 cDNA. This was performed uniformly for six chromosomes. This is the abbreviation for phenotype. This is how we are collecting data.

Researchers in the UK and Germany are doing the same thing. In the case of nematodes, they have collected RNAi phenotypes for 7000 to 8000, or perhaps over 10,000 genes. A large amount of this data is contained in a database called “WormBase” so it can be used. Phenotypes must be closely observed or cannot be observed unless conditions are changed. Duplication is therefore permitted, so researchers can try RNAi from various perspectives, study the phenotypes and accumulate the data.

We were particularly interested in the initial stage of generation, so we selected from among the gene groups manifested at an extremely early stage and did RNAi. When we did this, even though we usually do RNAi randomly, phenotypes appeared in at least 10 to 20% of the cases. If we establish criteria

limited to the early stages and then choose, nonphenotypes appear for 35%, so some sort of phenotype appears for 65%, so we were able to study the contents in detail.

When we did this, the strongest phenotype was arrested as a fertilized egg. If arrested immediately following fertilization, subsequent segmentation does not occur. As for other phenotypes, the nucleus behaves strangely. We were able to classify the various phenotypes by time sequence.

This gene for example becomes like this if impregnated. Although it may look the same, it is completely different. Here is the wild type, which is divided into large and small cells. This is called "asymmetrical division." By becoming asymmetrical, the front and rear can have different fates. This is because the contents change. These eggs divide equally. In this case, the difference between front and rear no longer exists. The egg subsequently divides equally, right in the middle. This one on the other hand differs in that the front divides quickly. In other words, in the case of this gene, there is an effect where the asymmetry disappears. Generation therefore does not progress and the cell cannot divide.

When this gene is knocked out, segmentation does no occur subsequent to fertilization. Although a small detail, meiotic division of the female nucleus does not occur. If you are wondering what this gene is, it is a homologous protesome, and is a sub-unit of protesome. We found that this sort of thing occurs. When we oppositely studied all other sub-units of the protesome, we found they exhibit the same phenotype. By classifying phenotypes in this way, we are oppositely able to further clarify the sub-unit structure of intracellular minute organs.

Last of all is computer modeling whereby people all over the world – this is recently referred to as "system biology" – many computer scientists are trying a single model case, and is a fascinating field. We are attempting to accurately apply the division pattern to a coordinate system. We first prepared some computer graphics. Here we have measured the behavior of the wild type as is and changed it to computer graphics. So this is cell contact, and locations and so on have all been measured.

Based on this, we must create a model that mimics and reproduces this sort of thing, so that is what we are preparing. Although still primitive, the membrane is expressed as the blue grainy mesh, and we are creating one that will be separated by the dynamic model. Up to now we have used four cells, if we turn this around – this spinning is extremely important – we are currently establishing parameters for skillfully turning this around.

So we have tried various things. Collecting all this data, we are creating a database called "NEXTDB," and are making it available to the public by this URL. We are also linked with WormBase so it can be viewed throughout the world.

This completes my presentation. There are many people that I must acknowledge. The nematode community has a tradition of publishing and sharing data, so I would like to cooperate with the community. This completes my talk on the present situation of functional genomics of nematodes. Thank you for listening.

Q&A

Blumberg: Thank you very much Prof. Kohara. We have time for one or two questions.

Sone: My name is Sone from National Institute for Environmental Studies. I do not know much about C. elegans. I was wondering if C. elegans are susceptible to cancer. Because of the short life span and small number of cells, I would think C. elegans would make a good model, a good strain for studying onset and differentiation.

Kohara: No matter where I go I am always asked if cancer occurs in nematodes. In that sense of the word, cancer does not occur, but there are cancer genes. If these genes mutate for example, many ovipositor and ovipositional organs are formed (multi-vulva). In some cases the number of times the cells divide increases dramatically, resulting in an abnormality. That is one type of cancer. I can explain all by cancer gene cascade, and I think that is one manifestation.

Inoue: If you have a question, please step up to the microphone.

Blumberg: OK, one last question.

Q: How do nematodes react to drugs, or rather, what is the toxicity?

Kohara: As for the reaction to drugs, nematodes are often used by [pharmaceuticals] companies for screening drugs. Nematodes are sometimes exterminated by neurotoxins and pesticides naturally, and research is consequently being conducted in that area. I am however not very familiar with such research.
Although it was not mentioned today, the group to which Dr. Arizono over there belongs is now conducting experiments involving gene variation using DNA chips on nematodes that have been exposed to environmental hormones. I think they will soon obtain various findings.