



DNA DISCOVERY

A group of researchers working at the Human Genome Project will be announcing soon that they made an astonishing scientific discovery: They believe so-called non-coding sequences (97%) in human DNA is no less than genetic code of an unknown extraterrestrial life form.

The non-coding sequences are common to all living organisms on Earth, from molds to fish to humans. In human DNA, they constitute larger part of the total genome, says Prof. Sam Chang, the group leader. Non-coding sequences, also known as "junk DNA", were discovered years ago, and their function remains

mystery. Unlike normal genes, which carry the information that intracellular machinery uses to synthesize proteins, enzymes and other chemicals produced by our bodies, non-coding sequences are never used for any purpose. They are never expressed, meaning that the information they carry is never read, no substance is synthesized and they have no function at all. We exist on only 3% of our DNA. The junk genes merely enjoy the ride with hard working active genes, passed from generation to generation. What are they? How come these idle genes are in our genome? Those were the question many scientists posed and failed to answer - until the breakthrough discovery by Prof. Sam Chang and his group.

Trying to understand the origins and meaning of junk DNA Prof. Chang realized that he first needs a definition of "junk". Is junk DNA really junk, (useless and meaningless) or it contains some information not claimed by the rest of DNA for whatever reason? He once mentioned the question to an acquaintance, Dr. Lipshutz, a young theoretical physicist turned Wall Street derivative securities specialist. "Easy," replied Lipshutz. "We'll run your sequence through the software I use to analyze market data, and it will show if your sequences are total garbage, "white noise", or

there is a message in there." This new breed of analysts with strong background in math, physics and statistics are getting more and more popular with Wall Street firms. They sift through gigabytes of market statistics, trying to uncover useful correlation between the various market indexes, and individual stocks.

Working evenings and weekends, Lipshutz managed to show that non-coding sequences are not all junk, they carry information. Combining massive database of the Human Genome Project with thousands of data files developed by geneticists all over the world Lipshutz calculated Kolmogorov entropy of the non-coding sequences and compared it with the entropy of regular, active genes. Kolmogorov entropy, introduced by the famous Russian mathematician half a century ago, was successfully used to quantify the level of randomness in various sequences, from time sequences of noise in radio lamps to sequences of letters in 19th century Russian poetry. By and large, the technique allows researchers to quantitatively compare various sequences and conclude which one carries more information than the other does. "To my surprise, the entropy of coding and non-coding DNA sequences was not that different", continues Lipshutz. "There was noise in

both but it was no junk at all. If the market data were that orderly, I would have already retired."

After a year of cooperation with Lipshutz, Chang was convinced, there is a hidden information in junk DNA. However, how could one understand its meaning if the information is never used? With active sequences you try to watch the cell and see what proteins are being made using the information. This wouldn't work with dormant genes. There will be experiment to test a hypothesis; one should rely on the power of his thought. Since there are letters, it should be tested in some old languages, perhaps Sumerian, Egyptian, Hebrew, and so on. Prof. Sam Chang solicited help from three specialists in the field, but none of them managed to find a solution. There were no cultural clues, no references to other known languages, the field was too alien for the linguists.

"I asked myself: who else can decipher a hidden message?" Chang continues. "Of course, cryptographers! In addition, I began talking with researchers at the National Security Agency. It took me few months to make them return my calls. Were they running background checks on me? Alternatively, were they too busy lobbying senators on retaining and strengthening

their authority to control exports of encryption technologies? Eventually, a junior fellow was assigned to answer my questions. He listened, requested my questions in writing and after another, few months turned me down. His message was polite but meant, "Go to hell with your crazy ideas. We are a serious agency, its National Security, dude. We are too busy."

Well, Sam, forget the Government, talk to the private sector. Therefore, I began approaching computer security consultants. They were genuinely interested, and a couple of them even began working on my project, but their enthusiasm always faded after a month. I kept calling them until one nice fellow told me: "I'd love to work on your project if I had more time. I am overbooked. Emissaries of major banks and Fortune 500 companies are begging me to plumb the holes in their networks. They pay me \$500 an hour. I can give you an educational discount, can you afford \$350?" Scrambling \$15/hr for post doctoral studies is a big deal in academia, \$350 sounded as something extraorbital." Eventually Prof. Chang was referred to Dr. Adnan Mussaelian, a talented cryptographer in the former Soviet republic of Armenia. Poor fellow barely survived on a \$15 a month salary and

occasional fees for tutoring children of Armenian nouveau riches. A \$10,000 research grant was a struck of luck, he began working like a beaver.

Adnan promptly confirmed the findings of his Wall Street predecessor: The entropy indicated tons of information almost in the clear, it was not too strong cryptographic system, it didn't appear to be a tough problem. Adnan began applying differential cryptoanalysis and similar standard cryptographic techniques.

He was two months in the project when he noticed that all non-coding sequences are usually preceded by one short DNA sequence. A very similar sequence usually followed the junk. These segments, known to biologists as alu sequences, were all over the whole human genome. Being non-coding, junk sequences themselves, alu are one of the most common genes of all.

Trained as a cryptographer and computer programmer, and having no knowledge of microbiology, Adnan approached the genetic code as of computer code. Dealing with 0, 1, 2, 3 (four bases of genetic code) instead of 0s and 1s of the binary code was a sort of nuisance, but the computer code was what he was analyzing and deciphering all his life. He was on familiar territory. The most common

symbol in the code that causes no action followed by a chunk of dormant code. What is that? Just playing with the analogy Adnan grabbed the source code of one of his programs and fed it into the program that calculates the statistics of symbols and short sequences, a tool often used in decoding messages. What was the most common symbol? Of course, it was "/", a symbol of comment! He took a Pascal code, and it were { and } ! Of course, the code between two slashes in C is never executed, and is never meant to be executed; it is not the code, it is the comment to the code!

Being unable to resist the temptation to further play with the analogy, Adnan began comparing statistical distributions of the comments in computer and genetic code. There must be a striking difference. This should show up in statistics. Nevertheless, statistically, junk DNA was not much different from active, coding sequences. To be sure, Adnan fed a program into the analyzer: surprisingly, the statistics of code and comments were almost the same. He looked into the source code and realized why: there were very few comments in between the slashes, it was mostly C code the author decided to exclude from execution, a common practice among programmers.

Adnan, religiously inclined

person, was thinking about the divine hand - but after analyzing the spaghetti code inside the sequences he convinced himself that whoever wrote the small code was not God. Who wrote the active, small coding part of human genetic code was not very well organized, he was a rather sloppy programmer. It looked like rather somebody from Microsoft, but at the time human genetic code was written, there was no Microsoft on Earth.

On Earth? It was like a lightning... Was the genetic code for all life on Earth written by an extraterrestrial programmer and then somehow deposited here, for execution? The idea was mad and frightening, and Adnan resisted it for days. Then he decided to proceed. If the non-coding sequences are parts of the program that were rejected or abandoned by the author, there is a way to make them work. The only thing one needs to do is to remove the symbols of comments and if the portion between the /*.....*/ symbols is a meaningful routine it may compile and execute! Following this line of thought, Adnan selected only those non-coding sequences that had exactly the same frequency distribution of symbols as the active genes. This procedure excluded the comments in Marcian or Q, whatever it was. He selected some 200 non-coding

sequences that most closely resembled real genes, stripped them of /*, //, and similar stuff and after few days of hesitation sent e-mail to his American boss, asking him to find a way to put them in E-coli or whatever host and make them work.

Chang did not replied for two weeks. "I thought I was fired", confessed Dr. Mussaelian. "With every day of his silence I more and more realized how crazy my idea was. Chang would conclude I was a schizophrenic and would terminate the contract. Chang finally responded and, to my surprise, he did not fire me. He had not bought my extraterrestrial theory but agreed to try to make my sequences work."

Biologists have attempted for years to make junk sequences express, without much success. Sometimes nothing turned out; sometimes it was junk again. It was not surprising. Grab an arbitrary portion of the excluded computer code and try to compile it. Most likely, it will fail. At best, it will produce bizarre results. Analyze the code carefully, fish out a whole function from the comments, and you may make it work. Because of careful Mussaelian's statistical analysis 4 of the 200 sequences he selected, began working, producing tiny amounts of a chemical compounds.

"I was anxiously awaiting the response from Chang," says Dr. Mussaelian. "Would it be a more or less normal protein or something out of ordinary? The answer was shocking: it was a substance, known to be produced by several types of leukemia in men and animals. Surprisingly, three other sequences also produced cancer-related chemicals. It no longer looked like a coincidence. When one awakens a viable dormant gene, it produces cancer-related proteins. Researchers began searching Human Genome Project databases for the four genes they isolated from junk DNA. Eventually, three of the four were found there, listed as active, non-junk genes. This was not a big surprise: since cancer tissues produce the protein, there must be somewhere a gene, which codes it! The surprise came later: In the active, non-junk portion of the code the gene in question (the researchers called it "jhlgl1", for junk human leukemia gene) was not preceded by the alu sequence, i.e. the /* symbol was missing. However, the closing */ symbol at the end of "jhlgl1" was there. This explained why "jhlgl1" was not expressed in the depth of the junk DNA but worked fine in the normal, active part of the genome. The one who wrote the basic genetic code for humans excluded portion of the big code by embracing

them in /*... */ but missed some of the opening /* symbol. His compiler seems to be garbage, too: a good compiler, even from terrestrial Microsoft, would most likely refuse to compile such program at all.

Prof. Sam Chang with his students began searching for genes associated with various cancers, and almost in all instances they discovered that those genes are followed by the alu sequence (i.e. protein as a comment closing symbol */), but never preceded by the comment opening /* gene! "This explains why diseases result in cell damage and their death, whereas cancers lead to cell reproduction and growth. Because only few fragments from the big code are expressed, they never lead to coherent growth. What we get with cancer, is expression of only few of genes alien to humans and symbiosis with some genes of bacterial parasites that lead to illogical, bizarre and apparently meaningless chunks of living cells. The chunks have its own veins, arteries, and its own immune system that vigorously resists all our anti-cancer drugs.

"Our hypothesis is that a higher extraterrestrial life form was engaged in creating new life and planting it on various planets. Earth is just one of them. Perhaps, after programming, our creators

grow us the same way we grow bacteria in Petri dishes. We can't know their motives - whether it was a scientific experiment, or a way of preparing new planets for colonization, or is it long time ongoing business of seedling life in the universe. If we think about it in our human terms, the extraterrestrial programmers were most probably working on one big code consisting of several projects, and the projects should have produced various life forms for various planets. They have been also trying various solutions. They wrote the big code, executed it, did not like some function, changed them or added new one, executed again, made more improvements, tried again and again. Of course, soon or later it was behind schedule. Few deadlines have already passed. Then the management began pressing for an immediate release. The programmers were ordered to cut all their idealistic plans for the future and concentrate now on one (Earth) project to meet the pressing deadline. Very likely in a rush, the programmers cut down drastically the big code and delivered basic program intended for Earth. However, at that time they were (perhaps) not quite certain which functions of the big code may be needed later and which not, so they kept them all there. Instead of cleaning the basic program by deleting

all the lines of the big code, they converted them into comments, and in the rush they missed few /* symbols in the comments here or there; thus presenting mankind with illogical growth of mass of cells we know as cancer."

There are three options to the problem. Either delete all the /* symbols and comments and clean this way the basic code, or add all the missing */ and avoid illogical mixing of the basic code with the big code. Alternatively, in the third option, remove all the / symbols and let work the basic code with the big code as a complete program. Unfortunately, none of these options are within our capacity. If we were able to efficiently insert genes into the chromosomes of living men, our breakthrough discovery would mean instant cure for all future cancer cases; at least from the programmer point of view. Theoretically, we can do it in a laboratory, but we have no practical means to implant the repaired DNA into living subjects. The mystery of "junk DNA" and cancer seems to be solved, but no quick cure shall be expected. The best thing we can do now is to try nourishing new, cancer-free line of humans with gradually debugged basic genetic code. That will take a long time. For us and our children, there is no hope on the horizon.

"However, from the

programmer's point of view, there is also positive outlook in it. What we see in our DNA is a program consisting of two versions, a big code and basic code. First fact is, the complete program was positively not written on Earth; that is now a verified fact. The second fact is, that genes by themselves are not enough to explain evolution; there must be something more in the game. What it is or where it is, we don't know. The third fact is, no creator of a new work, be it a composer, engineer or programmer, from Mars or Microsoft, will ever leave his work without the option for improvement or upgrade. Ingenious here is, that the upgrade is already enclosed - the "junk DNA" is nothing more than hidden and dormant upgrade of our basic code! We know for some time that certain cosmic rays have power to modify DNA. With this in mind, plausible solution is available. The extraterrestrial programmers may use just one flash of the right energy from somewhere in the Universe to instruct the basic code to remove all the /*...*/ symbols, fuse itself with the big code ("junk DNA") and jumpstart working of our whole DNA. That would change us forever, some of us within months, some of us within generations. The change would be not too much physical, (except no more cancers, diseases and short life), but it will catapult us intellectually.

Suddenly, we will be in time comparable to coexistence of Neanderthals with Cromagnons. The old will be replaced giving birth to a new cycle. The complete program is elegant, very clever self-organizing, auto-executing, auto-developing and auto-correcting software for a highly advanced biological computer with build-in connection to the ageless energy and wisdom of the Universe. Software wise, within us is either short and diseased life, or potential for a super-intelligent super-being with a long and healthy life. This triggers puzzling questions - was the reduction to the basic code done by sloppy programmers in a rush (as it appears to us), or was the disabling of the big code purposeful act which can be cancelled by a "remote control" whenever desired?"

Soon or later, we have to come to grips with the unbelievable notion that every life on Earth carries genetic code for his extraterrestrial cousin and that evolution is not what we think it is. This discovery may well shake the very roots of humanity - our beliefs in our concept of God and in our own power over our destiny. With the right paradigm, we may discover one day that all forms of life and the whole Universe is just one huge intellectual exercise in thoughts expressed mathematically, by Design, by Creator.

This article first appeared:

http://www.gewo.applet.cz/health/DNA_1e.htm