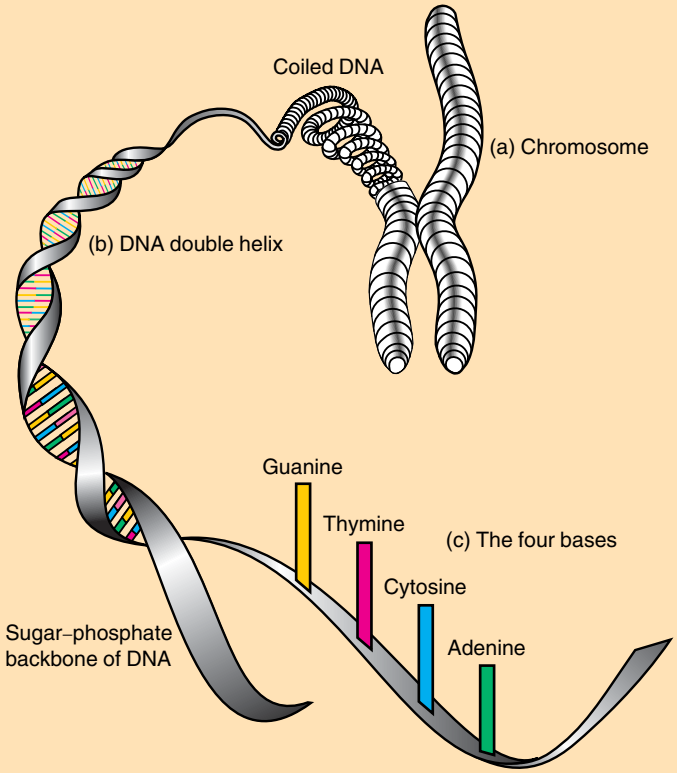
Chromosome 19 and Lawrence Livermore Form a Long-Lasting Bond

Livermore researchers join in the worldwide celebration of meeting the goals of the Human Genome Project.

THIS year marks the 50th anniversary of the discovery of DNA by researchers James Watson and Francis Crick. It seems historically fitting, then, that the complete sequence of all 23 human chromosomes was published earlier this year, thereby fulfilling the ultimate goal of the Human Genome Project, the most ambitious research effort in the history of the life sciences.

Biomedical scientists at Lawrence Livermore have played a prominent role in the Human Genome Project through their study of chromosome 19. Over the past two decades, dozens of Livermore researchers have discovered important new information about the 1,400 genes belonging to this chromosome. They determined the location of hundreds of genes (a process



Every cell in the human body (except red blood cells) contains 23 pairs of chromosomes. (a) Each chromosome is made up of a tightly coiled strand of DNA. (b) DNA's uncoiled state reveals its familiar double helix shape. If DNA is pictured as a twisted ladder, its sides, made of sugar and phosphate molecules, are connected by (c) rungs made of chemicals called bases. DNA has four bases—adenine, thymine, guanine, and cytosine—that form interlocking pairs. The order of the bases along the length of the ladder is the DNA sequence.

called mapping) on chromosome 19, discovered the function of many of its genes, and began the enormous task of sequencing the chromosome, that is, determining the exact order of its DNA base pairs. Later, they participated in the complete sequencing effort as part of the Department of Energy's Joint Genome Institute (JGI) in Walnut Creek, California.

With the sequencing of chromosome 19 complete, Livermore scientists are helping to shape a new era in which the function of all genes and the proteins they produce are understood and medical professionals will be able to diagnose, treat, and perhaps cure the approximately 5,000 known hereditary diseases. To accomplish these ambitious goals, the researchers are comparing the human genome to that of other organisms such as the mouse, rat, chicken, and pufferfish. They are also studying the complex mechanisms that govern how some genes regulate the actions of others. Finally, they are building new computer tools to help make sense of the human genome.

Draft Sequence in 2001

It was barely two years ago, in April 2001, that DOE announced the draft decoding of chromosomes 5, 16, and 19 by JGI. Livermore biomedical scientist Lisa Stubbs notes that the historic milestone was only a first step because the draft sequence contained gaps and errors. Nevertheless, because the draft covered 90 percent of the human genome, it allowed scientists to identify thousands of genes, some of them responsible for inherited diseases.

The final sequencing steps, done at 20 genome centers worldwide, filled most of the gaps in the sequence and increased the overall accuracy to 99.99 percent, or one error per 10,000 bases. "It's really good to have the sequencing finished," says Stubbs. She says the final sequencing steps performed at JGI and Stanford University show that the draft sequence "pretty much got it right."

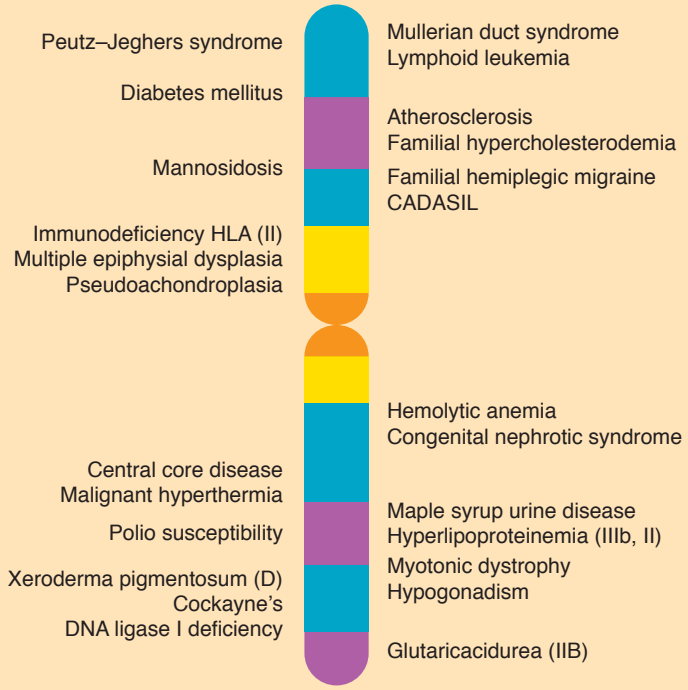
All chromosomes are numbered according to their length, with chromosome 1 being the longest. Chromosome 19 is one of the smallest and, with about 65 million bases, most gene-dense of the human chromosomes. It is home to the genes that are linked to lymphoid leukemia, myotonic dystrophy, diabetes mellitus, atherosclerosis, and a susceptibility to polio along with dozens of other heritable conditions.

Livermore's chromosome 19 research was a natural outgrowth of the work in its biomedical department, which was chartered in 1963 to study the radiation dose to humans from isotopes in the environment. Radiation was known to cause damage in chromosomes, and scientists believed that a useful way to learn about the effects of radiation and other environmental toxins was to study DNA directly. (See *S&TR*, November 2002, pp. 22–30.)

Early Start on Chromosome 19

Livermore researchers chose chromosome 19 to study because, of all 23 human chromosomes, it has the highest concentration of guanine–cytosine base pairs, long thought to imply a higher concentration of genes. That hunch proved to be correct. "The density of genes on chromosome 19 is one of the highest of all chromosomes," says Livermore biomedical scientist Laurie Gordon.

One early Livermore research project examined three genes on chromosome 19 that are involved in the repair of DNA damaged by radiation and environmental pollutants. DNA repair genes produce proteins that "cruise" the



Dozens of genes associated with heritable diseases are located on chromosome 19.

length of DNA, removing unwanted proteins and looking for any mistakes that might disrupt the smooth functioning of a cell. These repair gene studies, which still continue, may lead to insights about the development of cancers, many of which are caused by defects in DNA repair pathways. Another Livermore project studied a family of about 60 genes on chromosome 19 involved in detoxifying and excreting chemicals foreign to the human body.

Livermore research on chromosome 19 accelerated when, in 1986, DOE launched a major initiative to completely decipher the human genetic code. Soon, Livermore researchers were studying all of chromosome 19, with Lawrence Berkeley researchers focusing on chromosome 5 and Los Alamos scientists on chromosome 16. In 1990, DOE joined the National Institutes of Health to launch the U.S. portion of the Human Genome Project with the goal to discover all the human genes and to determine the complete sequence of the genome's 3 billion DNA base pairs. The project soon drew additional collaborators worldwide.

Biomedical scientist Linda Ashworth worked on chromosome 19 for 14 years before retiring in 2001. "When I first got involved in the mid-1980s, genome science was a very small field," she recalls. Ashworth, Anne Olsen, and others started mapping chromosome 19 in the mid-1980s to understand the location of hundreds of genes that were then known to reside on the chromosome and to prepare for the sequencing effort that lay ahead.

Mapping Effort Not Easy

The tedious process, which Ashworth describes as "putting Humpty-Dumpty together again," involved breaking up thousands of chromosome 19 molecules into small pieces of the same size; producing exact copies, called clones, of each piece with bacterial colonies; and then fitting the pieces together in the correct order. The technique was the only option available at the time because scientists were (and still are) unable to sequence an entire chromosome from end to end. "Our goal was to create a high-resolution sequence-ready map," says Olsen, who also retired recently.

Olsen says that chromosome 19 was not easy to map. It contains a large number of repetitive sequences that are interspersed throughout its DNA. Another factor, which complicated accurate sequencing later on, is that the chromosome is more tightly bonded than other chromosomes because of its high instances of guanine–cytosine bonds. As a result, a single strand of the double-stranded molecule can loop back on itself to form confusing secondary structures.

During the mapping effort, Livermore researchers tapped the latest tools and techniques in the fledgling biomedical industry, such as automated pipettes, and invented a few of their own. One important technique, developed at the Laboratory to locate short pieces of DNA and establish their relative position on an individual chromosome, is called fluorescence in

(a) Robots have revolutionized many formerly labor-intensive activities at the Joint Genome Institute. A robot selects bacterial colonies with large amounts of cloned human DNA and transfers to machines that will sequence the DNA. (b) These high-speed DNA sequencers at the Joint Genome Institute can sequence 2 billion base pairs per month.

situ hybridization (FISH). Researchers improved the resolution of FISH by using hamster eggs fused with individual human sperm, which caused the sperm DNA to extend in length. This extension allowed investigators to see the molecule in much greater detail than was previously possible.

Sequencing Begins

In the late 1980s, armed with their map of known genes along the chromosome, Livermore scientists started sequencing the 65 million base pairs of chromosome 19. Ashworth recalls, "When we started sequencing, it took about 14 hours to sequence 400 bases, but that was considered state of the art."

As the mapping efforts continued, other Livermore researchers discovered more about the location and function of chromosome 19's genes. In 1992, Livermore researchers, collaborating with colleagues in Canada and Europe, discovered the genetic defect that causes myotonic dystrophy, the most common form of muscular dystrophy.

When the Joint Genome Institute was established in 1997 as a collaboration between Livermore, Los Alamos, and Lawrence Berkeley national laboratories, Livermore researchers sequenced 200 million raw base pairs in one year—that is, they gave the DNA its first rough reading. JGI can currently complete 200 million raw bases in less than 3 days.

With JGI's production facility in full swing by mid-1999, sequencing took on an industrial character by centralizing and largely automating the effort that was being done individually at the three laboratories. JGI personnel, including a team led by biomedical scientist Susan Lucas, worked to complete the sequencing of chromosomes 5, 16, and 19.

JGI transferred the final sequencing work to a group at Stanford University

The Joint Genome Institute: From Virtual Facility to Gene Research Powerhouse

Located in Walnut Creek, California, the Department of Energy's Joint Genome Institute (JGI) is one of the largest publicly funded genome sequencing centers in the world. The institute was founded as a virtual entity on January 1, 1997, as a collaboration between Lawrence Livermore, Lawrence Berkeley, and Los Alamos national laboratories. Livermore scientists initially mapped and began to sequence chromosome 19, while Los Alamos scientists worked on chromosome 16, and Lawrence Berkeley worked on chromosome 5 before joining forces through the JGI. (See *S&TR*, April 2000, pp. 4–11.)

The main work of the JGI is done at its 5,600-square-meter Production Genomic Facility (PGF). Secretary of Energy Bill Richardson was keynote speaker at the April 19, 1999, formal PGF dedication. Its staff of about 150 includes 40 Lawrence Livermore researchers.

The PGF uses an automated process during which the samples pass through capillaries as a laser scans them. After DNA bases are read, computers reassemble the overlapping fragments into long, continuous stretches of sequenced DNA, which are analyzed for errors, gene-coding regions, and other characteristics. This process is repeated many times for all of the sections of DNA that make up a genome. The front end of the operation, where the pieces of DNA are cut, involves the most skilled handwork. Virtually all other facets of the process have been automated.

Livermore biomedical scientist Elbert Branscomb served as JGI's first director. Edward Rubin, an internationally known geneticist and medical researcher, was named the current director in January 2003. Funding is provided mainly by the Office of Biological and Environmental Research in DOE's Office of Science, with additional funding from the National Science Foundation, the U.S. Department of Agriculture, and other agencies.

JGI's initial goal was completing the DNA sequencing of chromosomes 5, 16, and 19, which together constitute 11 percent of the human genome. In April 2001, JGI announced the completion of the draft of JGI's three chromosomes. JGI was the first large genome center to make such an announcement, several months ahead of schedule.

After completing the final sequencing, JGI researchers sent their data to a team at Stanford University for "finishing," that is, closing the gaps and resolving any discrepancies in the draft sequence. To be considered finished, the sequence must be completely contiguous, be confirmed by at least two templates, have no gaps, and have a final estimated error rate of less than 1 out of 10,000 bases.

The center has taken advantage of innovations and breakthroughs in the bioresearch field, which have resulted in remarkable increases in the amount of DNA that is sequenced. Since 1999, the JGI has increased its production rates more than 20-fold to sequencing about 35 million bases per day.

JGI is in the process of becoming a more research-oriented facility. It has established whole genome sequencing programs that include vertebrates, fungi, plants, and bacteria.

for what is known as finishing. This process improves accuracy and closes small gaps in the known sequence. Stubbs notes that because the human genome will be used as a reference for all scientists, it is essential that it be as accurate as possible.

Entering a New Era

With the final decoding of the human genome, scientists are entering the postgenomic era. The new focus, says Stubbs, is on understanding the function, regulation, and evolution of genes. After a gene is precisely located on a chromosome and sequenced, researchers can easily predict the primary structure of the protein that the gene encodes. But in only a few cases do scientists have an idea what that protein does. Sometimes basic function can be deduced by similarity to other known proteins. For example, protein structure may suggest a role in detoxification or as a structural component.

"We know what only a handful of genes are doing in the organism," says Stubbs. "In most cases genes are like black boxes. What kinds of toxins do they metabolize? In which cells, and when do the cells require them? A small number of genes do not encode proteins at all but are suspected of producing RNA products that help regulate other genes."

Stubbs says that scientists also need to explore the largely uncharted noncoding region, sometimes called junk DNA, which makes up about 95 percent of human DNA. Although junk DNA looks like nonsense coding, it may have regulatory functions, or it might be necessary for the structural integrity of the DNA double helix. Gordon believes that junk DNA may yield surprising functions. Curiously, yeast and bacteria do not have junk DNA.

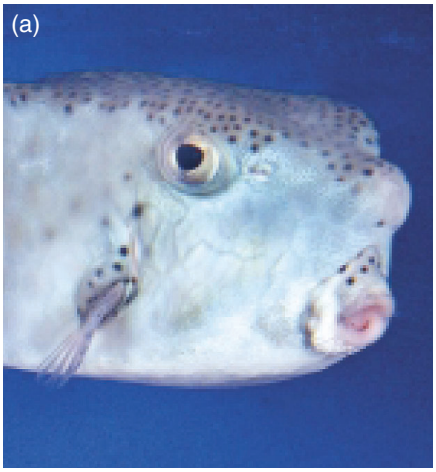
Many scientists like Stubbs and Gordon are interested in how genes are regulated. Two genes may have almost identical sequences, yet one may be active only in skin and the other only in bones. What determines the tissue-specific activity are DNA elements linked to the genes. These so-called regulatory elements serve as docking sites for proteins that determine the "on–off" state of every gene. Understanding gene regulation is important to a broad range of applications, ranging from human susceptibility to disease to managing microbes in the environment.

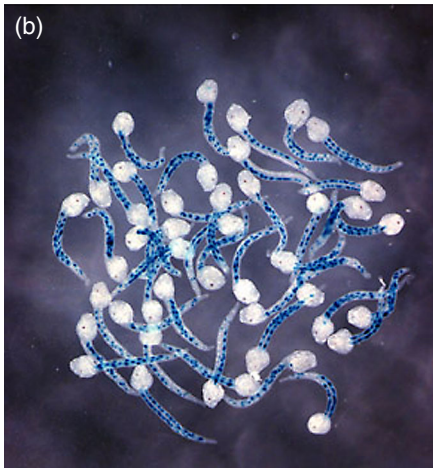
Livermore researchers have discovered possible regulatory sequences for genes throughout chromosome 19. "The complexity involved in gene regulation becomes exponential the more we look into it." says Gordon. " The cascade of reactions is fascinating."

Comparing Genomes

To gain further understanding of the human genome, many scientists are looking to comparative genomics. This relatively new field analyzes and compares the genetic material of different species to study evolution and gene function. A key goal is to find the five percent of most complex genomes that serve critical roles in cell development, maintenance, and health.

A pioneering study led by Livermore researchers showed the high degree of similarity between the mouse and human genomes.





In 2001, the JGI led a consortium that sequenced (a) the Japanese pufferfish, *Fugu rubripes*, and (b) the sea squirt, *Ciona intestinalis*. The pufferfish is the first vertebrate genome after human to be draft sequenced. The sea squirt is a primitive chordate and has a small genome.

Comparative genomics focuses on identifying DNA sequences that are shared between two or more diverse species. These conserved elements (that is, those that aren't reinvented for each species) include genes that produce key proteins and enzymes and the genes that establish basic features of the organism such as body shape during development. "If you really want to discover the genes that are most critical to basic life, you have to look at other genomes," says Stubbs.

Livermore scientists are discovering that humans share a surprising number of genes with mice, fish, chickens, and even primitive bacteria. A Livermore–JGI team led by Stubbs compared human chromosome 19 with similar sections of mouse DNA and detailed its findings in the July 2001 issue of *Science*. The article, the first major example of the power of large-scale genomic comparisons, described clues to the mechanisms of gene evolution in mammals. The research was also helpful because the mouse is often used as a model for studying diseases and testing medicines. (See *S&TR*, May 2001, pp. 12–20.)

The researchers found that functional counterparts of about 90 percent of the human genes in chromosome 19 are also located in similar sections of mouse DNA. However, against this backdrop of amazingly high similarity, they also found some significant differences. Most of the differences are due to the active copying of certain genes over 80 million years of evolution. Both rodents and primates have copied genes from the earliest chordates (animals with backbones), but not the same genes. As a result, rodents have multiple copies of some genes that are found only once in human DNA, and vice versa.

"Duplicated genes will often specialize, with each copy taking on parts of the original gene's function," says Stubbs. "For example, one copy may take over duties in the liver, while the other one specializes for work in the brain. This specialization is a way to build more complexity over time. As the genes mutate and additional duplications take place, completely new functions may arise. Gene duplication is therefore a major source of genetic variation, the critical fodder for evolution."

One group of genes that has duplicated especially frequently is called zinc finger genes, which produce proteins that bind to elements allowing them to regulate the activity levels of other genes. A significant fraction of the roughly 800 human zinc finger genes are found on chromosome 19.

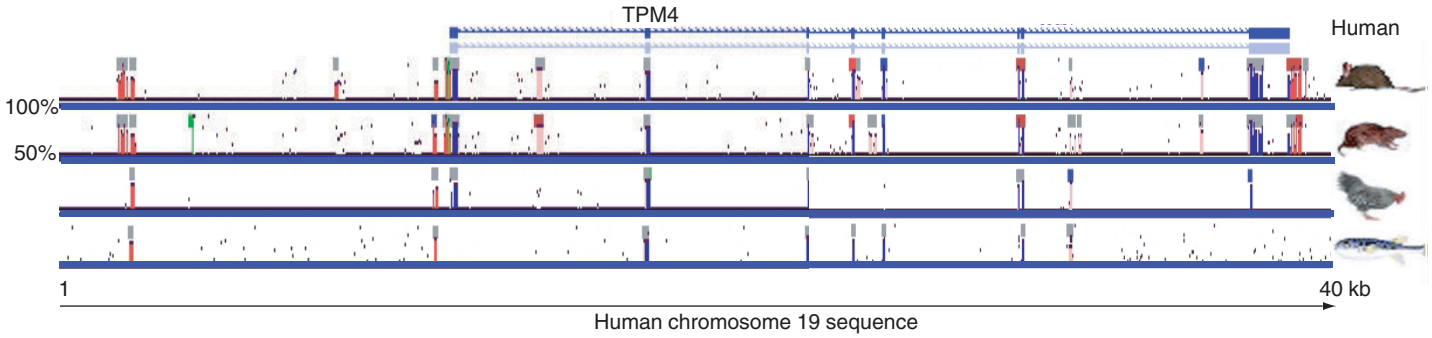
Gaining Insight with Pufferfish

Stubbs says that scientists need to look at genomes of organisms farther away evolutionarily from humans to gain further insight into the human genome. In early 2001, JGI led a consortium that sequenced the Japanese pufferfish, *Fugu rubripes*. It was the first vertebrate genome to be sequenced after the human genome. The pufferfish genome is known as the *Reader's Digest* version of the human genome because it is one-eighth the size. Nevertheless, it contains a gene set that is highly similar to the human genome, and its sequence is helping scientists to further identify regulatory genes in humans.

Later in 2001, JGI sequenced the genome of the sea squirt, *Ciona intestinalis*. This organism, a primitive chordate, has a small genome (165 billion base pairs) and therefore holds important information about how genes evolved over time.

Stubbs, Gordon, and other Livermore researchers have begun analyzing portions of the chicken genome and comparing it to the human genome. They chose the chicken because it sits on the evolutionary tree between the mouse and the pufferfish. Gordon points out that Livermore

Livermore researchers have begun analyzing the chicken genome and comparing it to the human genome because the chicken sits on the evolutionary tree between the mouse and the pufferfish.



A linear map (top) of a 40-kilobase (kb) region of human chromosome 19 containing the tropomysin 4 (TPM4) gene. Nucleotide positions 1 to 40,000 are represented by the horizontal axis of each panel. The seven segments, or exons, of the gene sequence are shown by boxes joined by a blue line. The four panels below the TPM4 sequence summarize the results of aligning this human sequence with similar regions of the genome of the mouse, rat, chicken, and pufferfish. Wherever matches greater than 50-percent identity are found between the human and other sequences, a dot is plotted in each panel at a height that corresponds to the percentage. A series of clustered dots may coalesce into a larger, evolutionarily conserved region (ECR). Blue ECRs correspond to gene exons; red ECRs correspond to nongene regions. The nongene regions are likely to represent regulatory sequences that control gene expression. TPM4 is well conserved between human and mouse, rat, birds, and fish. The figure is taken from the ECR browser designed by collaborator Ivan Ovcharenko of Lawrence Berkeley National Laboratory (nemo.lbl.gov/~ovcharen/).

researchers do not work among chicken coops. Instead, they use chicken genes that are available as clones produced by bacterial colonies. The chicken study, as with the mouse effort, largely involves complex computational analyses of the different genomes.

Indeed, computational genomics has become essential. These powerful computer tools allow users to ask complex questions and extract meaningful answers rapidly from massive amounts of genome data. Some of the tools have been developed by Livermore researchers such as bioinformaticist Paramvir Dehal, who is helping to compare different species' DNA, and computer scientist Art Kobayashi, who is developing new ways to analyze larger pieces of sequenced DNA than is now possible.

Smart Choice

Many new projects to better understand the human genome are just beginning. Gordon says that it's quite likely that more human genes will be discovered and their functions elucidated. "We have years and years of work ahead of us," she says. Increasingly, that work is done as collaborative efforts with other institutions.

Gordon notes that another large task is understanding the many variations among genes. Every human has a slightly different genome than everyone else. Just as certain genes determine the blood types A, B, AB, and O, other genes are responsible for proteins that are slightly different from each other.

In retrospect, says Stubbs, selecting chromosome 19 was an excellent choice for Livermore bioresearchers. It is small, therefore manageable, and twice as gene-dense as many other chromosomes. Working on the chromosome led to Livermore's present stature as a world leader in genomics, bioinformatics, and comparative genomics, and to its full participation in JGI. Knowledge about chromosome 19 and other chromosomes will speed the understanding of how genes influence disease development, contribute to the discovery of new treatments, illuminate how species evolved, and help scientists form a molecular understanding of life.

Many researchers, even those who have retired, still have strong feelings about chromosome 19. "Those of us who have worked with chromosome 19 for so many years have an emotional attachment to it," says Olsen.

—Arnie Heller

Key Words: chromosome 19, comparative genomics, DNA, gene mapping, gene sequencing, Human Genome Project, Joint Genome Institute (JGI), pufferfish, regulatory genomics, sea squirt.

For further information contact Lisa Stubbs (925) 422-8473 (stubbs5@llnl.gov).