IMAGING AND VISUALIZATION

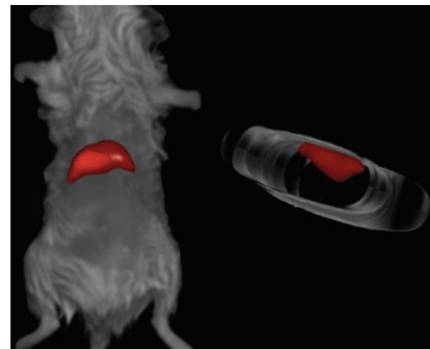
Fluorescent proteins: into the infrared

An engineered infrared fluorescent protein is the first member of a new class of genetically encodable probes, with special advantages over visible-wavelength fluorescent proteins for *in vivo* imaging.

Fluorescent proteins have come along way since the discovery of GFP from the jellyfish *Aequorea victoria* in the early 1960s. Today, both natural and engineered fluorescent proteins come in a veritable rainbow of colors, largely thanks to pioneering work done in Roger Tsien's laboratory at the University of California at San Diego, work that was recently recognized with part of the 2008 Nobel Prize in Chemistry.

In combination with fluorescence microscopy, genetically encodable fluorescent proteins are used to highlight various cell structures and monitor cellular processes in technicolored splendor. Fluorescent proteins are also useful for tracking cell populations, such as tumor cells, *in vivo*. Thus far, the best tools for imaging in live, whole animals have been far-red fluorescent proteins because both autofluorescence and light absorption by the animal's tissue are minimized at the far-red end of the visible spectrum. However, to produce far-red fluorescence, these proteins need to be excited by light at around or below 600 nanometers, which is substantially attenuated by hemoglobin before reaching deeply buried tissues. Despite continued improvement, so far there has been no success in developing a fluorescent protein with an excitation maximum around 700 nm, the absorption minimum of hemoglobin.

The Tsien lab recently took a new approach to develop a better fluorescent protein for *in vivo* imaging. Rather than begin with a red fluorescent protein scaffold and engineer it to make it brighter or shift the excitation further to the far-red, they decided to start fresh with a new scaffold to develop an infrared-emitting fluorescent protein. The researchers began with a bacterial phytochrome scaffold from *Deinococcus radiodurans*, which absorbs light at around 700 nm. This phytochrome binds a cofactor called biliverdin, which



Mouse liver in a living mouse labeled with an infrared fluorescent protein (represented by red pseudocolor) delivered via an adenoviral vector. A tilted view of the liver is shown on the right. Figure courtesy of Xiaokun Shu and Roger Tsien.

serves as a chromophore and just so happens to be found naturally in all animals. "The truncated phytochrome crystal structure was available, which facilitated protein engineering," explains Xiaokun Shu, a postdoc and the first author on this work. "Our goal with structure-based engineering was to rigidify

© 2009 Nature America, Inc. All rights reserved.

BIOINFORMATICS

GOING WITH THE SKEWED FLOW

Computational and experimental biologists teamed up to develop a new software tool to analyze the rich data generated by new and powerful flow cytometers.

Imagine watching the flow of the East Australian Current; how would you describe all the fish swimming in front of you or find a rare species? Biologists face a similar situation when they look at blood flow. But unlike the tropical fish that come in various sizes, shapes and colors, most blood cells can only be distinguished by their surface protein markers.

Immunologists have used flow-cytometry technology for studying blood cells for decades. A flow cytometer analyzes the cells flowing through a fluorescence detector one by one and records the surface-marker profile of each cell. Recent advances in new fluorescent dyes and detection techniques have allowed simultaneous detection of close to 20 different markers. However, new computational methods are needed to properly analyze these rich data. At the Broad Institute of the Massachusetts Institute of Technology, Jill Mesirov, a computational biologist, recently collaborated with two experimental biology groups led by Philip De Jager and David Hafler to develop a software tool called flow analysis with automated multivariate estimation (FLAME), for such flow-cytometric data analysis.

"The traditional way is to look at the data two or three markers at a time and manually identify cell subpopulations, but when you have 12 different markers, you can't look at 12-dimensional space," says Mesirov. The current approach is inefficient and subjective. Worse, explains De Jager, "if you don't know what you are looking for, you get more and more limited in the successive two- or three-dimensional projections and fail to recognize the architecture of the whole cell population."

Two important characteristics of flow-cytometric data, asymmetric distribution and outliers, complicate the analysis. "The histograms that come out of a flow cytometer always have tails, which are not captured very well if one assumes symmetric distribution," says De Jager. "And the outlier can be also quite important," adds Mesirov. Others have used symmetric modeling or data transformation to analyze high-dimensional flow-cytometric data. However, symmetric modeling does not fit the flow-cytometric data, and "the problem of data transformation is that very different asymmetric distributions can, after transformation, yield the same Gaussian distribution," explains Mesirov.

To model robustly against such asymmetry and outliers for precise identification of subpopulations from the high-dimensional data, the authors used a non-Gaussian statistical model based on the multivariate skewed *t* distribution. "Mathematically it was obvious that a new modeling approach was needed" for flow-cytometric data, says Mesirov. A postdoc in her laboratory, Saumyadipta Pyne, designed the high-dimensional mixture model and then worked with another postdoc, Kui Wang, in Geoffrey McLachlan's lab at the University

the bound biliverdin in order to decrease its nonradiative decay, so that absorbed 700-nm photons can be emitted through fluorescence."

Using saturation mutagenesis and DNA shuffling to vary the protein residues around the biliverdin chromophore, Shu, Tsien and their colleagues evolved a bright, photostable and monomeric infrared fluorescent protein mutant, with an excitation maximum of 684 nm and an emission maximum of 708 nm. When they expressed the mutant in the livers of mice via an adenoviral vector that specifically targets the liver, they observed strong infrared fluorescence.

Besides *in vivo* imaging applications, infrared fluorescent proteins will also be useful for cell-based imaging. "Infrared fluorescent proteins provide another color well-separated from those of existing fluorescent proteins," notes Shu. Furthermore, cell autofluorescence in the infrared region of the spectrum is nearly nonexistent, thus providing clearer images. Additionally, infrared fluorescent proteins may find application in fluorescence resonance energy transfer (FRET).

Infrared fluorescent proteins also need not be limited to being engineered from the phytochrome scaffold from *D. radiodurans*. There are many bacterial phytochrome sequences available, with absorption maxima varying from 650 to 750 nm. "Therefore," says Shu, "they are potential candidates for engineering blue- or red-shifted infrared fluorescent proteins, which then may be used for multicolor imaging and as FRET donors or acceptors for *in vivo* FRET imaging."

Allison Doerr

RESEARCH PAPERS

Shu, X. *et al.* Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome. *Science* **324**, 804–807 (2009).

© 2009 Nature America, Inc. All rights reserved.

of Queensland in Australia to create efficient software code to carry out the statistical analysis. They incorporated the code into a user-friendly interface, the FLAME software.

The Mesirov and De Jager groups tested FLAME with data sets derived from peripheral mononuclear blood cells. Using two-step clustering with multiple surface markers, they isolated an important regulatory T-cell population that consists of only 0.81% of the total number of blood cells. However, FLAME does more than identify subpopulations. Aided by another mathematical approach called bipartite graph matching, FLAME can be used to detect changes to the surface marker profile of specific cell subpopulations in distinct states. The researchers demonstrated this by matching cell populations before and after T-cell stimulation and visualizing phosphorylation and other marker shifts.

"The FLAME output is much richer than the old way of looking at flow-cytometric data," said De Jager. As a neurologist, he plans to use FLAME to discover biomarkers for neurodegenerative diseases, such as multiple sclerosis. Meanwhile, Mesirov's group is tackling more complex data, such as the nonconvex distributions, to examine biological systems by mathematical description. Also, by adding FLAME to the Gene Pattern genomic software package developed by the Mesirov group, researchers will be able to compare flow-cytometric data to other high-throughput data, according to Mesirov.

Wayne Peng

RESEARCH PAPERS

Pyne, S. *et al.* Automated high-dimensional flow cytometric data analysis. *Proc. Natl. Acad. Sci. USA* **106**, 8519–8524 (2009).

NEWS IN BRIEF

GENE TRANSFER

Transgenic marmosets express EGFP

Nonhuman primate models of human disease would be extremely valuable for biomedical research. By injecting marmoset embryos with a lentiviral vector, Sasaki *et al.* have now generated transgenic marmosets that express enhanced GFP (EGFP). Besides expressing EGFP in somatic cells, two of the transgenic marmosets expressed EGFP in germ cells, and one transgenic male passed the 'green' gene onto his healthy offspring.

Sasaki, E. *et al. Nature* **459**, 523–527 (2009).

GENOMICS

Sequencing pools of tens of thousands

In current multiplexing strategies for high-throughput sequencing, a molecular barcode is appended to each sample, allowing the parallel analysis of dozens of specimens. Erlich *et al.* now expand this multiplexing capability by three orders of magnitude with a smart pooling approach. They assign barcodes to pools, rather than to individual specimens, and identify each sample by decoding the pooling pattern. This approach allowed them to sequence and decode a pool of 20,000 different artificial microRNAs.

Erlich, Y. *et al. Genome Res.* advance online publication (15 May 2009).

CHEMICAL BIOLOGY

Unnatural DNA bases put to the test

The creation of unnatural but functional DNA base pairs has been a long-standing goal in chemical biology. Several unnatural base pairs have been developed and tested *in vitro*, where they have been accepted by DNA polymerases. Delaney *et al.* now describe the first test of unnatural base pairs in living cells. They demonstrated that two size-expanded base pairs were efficiently read by the replication machinery in *Escherichia coli*.

Delaney, J.C. *et al. Angew. Chem. Int. Ed.* **48**, 4524–4527 (2009).

BIOINFORMATICS

A tool to find regulatory elements in RNA

Regulation of gene expression occurs at many different levels, including RNA stability. Foat and Stormo now present a computational approach to identify secondary structure–defined *cis*-regulatory elements (SCREs) in mRNAs, by modeling their effect on mRNA levels as measured by microarrays. They applied their algorithm, named StructRED, to all stem-loops 9–12 nucleotides in length and recovered the known binding specificities of two RNA binding proteins.

Foat, B.C. & Stormo, G.D. *Mol. Sys. Biol.* **5**, 268 (2009).

NANOTECHNOLOGY

Quantum dots without blinking

Wang *et al.* describe the synthesis of quantum dots that exhibit continuous photoluminescence. Other groups have developed various strategies to suppress blinking, but none have completely eliminated blinking. Wang *et al.* achieved this by making core nanocrystals of CdZnSe coated with a semiconductor shell of ZnSe. Though these nanocrystals have not yet been tested in biological applications, the fact that they do not blink makes them promising for use in single-molecule experiments.

Wang, X. *et al. Nature* **459**, 686–689 (2009).