second movies, to brain activity patterns. Their dark, blurry movie reconstructions are an average of the top 100 short movie clips with a predicted brain pattern that fits best to the actual brain activity patterns.

Naselaris cautions that they can’t recreate dreams or other mind’s-eye images with this model. Because they focused on an early visual brain region, their model “has everything to do with the light that is hitting your eye,” Naselaris says.

Yet Gallant and his group’s strategy is powerful for a number of reasons. With it, they can make surprisingly accurate predictions about images that the model has not seen before. In addition, their strategy of first creating an encoding model before decoding gives researchers a method for testing theories about how the brain processes information. For other areas of the brain, the strategy will likely have to be modified. “The brain is a pretty complicated place,” Gallant says, “so there is no one grand approach that will solve everything. Instead, there are thousands of neuroscientists using thousands of different approaches to try to move ahead.”

The brain is a pretty complicated place,” Gallant says, “so there is no one grand approach that will solve everything. Instead, there are thousands of neuroscientists using thousands of different approaches to try to move ahead.”

PERSONALIZED CANCER TREATMENT: Seeking Cures Through Computation

By Kristin Sainani, PhD

Personalized cancer therapy is now a reality. A handful of tumor-classifying tests and targeted drugs are in widespread clinical use; and early attempts are underway to match high-risk cancer patients to experimental drugs based on genetic testing of their tumors.

But progress has been incremental and successes have been measured. Cancer is complex and insidious; knock out one bad player with a drug and the system evolves resistance. Patients may live longer, but still die of their disease. To take personalized cancer medicine to the next level—to achieve cures—computational approaches are needed. “We are at a crossroads where it’s becoming increasingly difficult to do anything of value without a heavy element of computation,” says Andrea Califano, PhD, professor of biomedical informatics at Columbia University and director of the Columbia Initiative in Systems Biology.

Bioinformatics and computing are helping to make advances on several fronts, including: cataloging the full spectrum of genomic defects in cancer; identifying the defects that drive malignancy; efficiently translating these discoveries to patient care; and improving the tools that are already in clinical use.

Mapping the Landscape

Several cancer genome initiatives are cataloging the array of molecular defects that define different cancers and cancer subtypes. The NIH’s The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) have already collected multiple layers of data—including sequencing, mutational,
copy number alteration, DNA methylation, microRNA, and gene expression data—on an unprecedented number of tumors.

“These different layers will be mapped out, overlaid, and integrated, so we can see the complete genomic picture of the tumor,” says Douglas A. Levine, MD, head of the Gynecology Research Laboratory at the Memorial Sloan-Kettering Cancer Center and a TCGA investigator. This is the first time we’ve had all these different data types on exactly the same samples, Califano says. “And simply amazing findings are coming out.”

The TCGA reported results for its second complete genome—for high-grade serous ovarian cancer (a common and aggressive form of this cancer)—in Nature in June 2011. The project analyzed data from 489 tumors, including the complete exome sequences of 316, the most ever reported to date for any solid tumor, Levine says.

Among the most exciting findings, about half the patients had defects—inherited or acquired mutations or epigenetic silencing—in the tumor suppressor genes BRCA1 and BRCA2 or in related DNA repair genes. These tumors might respond to PARP inhibitors, which improve survival in women with inherited BRCA1 and BRCA2 mutations, Levine says.

Also, seven percent of tumors had homozygous deletions in the tumor suppressor gene PTEN, a defect not previously reported in this subtype of ovarian cancer. Levine and others have already begun clinical trials to treat this subset of patients with a new class of drugs called PI3 kinase inhibitors, which target the PI3K/AKT/mTOR pathway that PTEN regulates.

“All these things need to be tested. But we now have the landscape and the roadmap laid out,” Levine says. “Molecular medicine is not really being used at all today in ovarian cancer. I hope it can be used in the near future to make better treatment decisions.”

Identifying the Driving Defects

Sequencing cancer genomes is only the first step in understanding this disease; the next step is to sort out which genetic changes are driving the cancer—and thus will make robust biomarkers and drug targets—and which are merely incidental. This is where systems biology comes in handy, Califano says. “The idea is to computationally interrogate regulatory networks of the cancer cell to find out what are the genes that are actually causally related to the presentation of a specific tumor phenotype,” Califano says.

For example, in a 2010 paper in Nature, Califano’s team used network analysis to identify two master regulators of a particularly aggressive subtype of glioma brain cancer. These two transcription factors (C/EBPβ and Stat-3) don’t appear in the gene expression signature for this subtype, as they are about the 500th and 1300th most differentially expressed, Califano says. “But, if you look at them with these network analyses, they stand out as being the most significant genes in terms of their activity in regulating the signature.” Inactivating these genes in mouse xenografts blocked tumor development or reduced malignancy.

In an October 2011 paper in Cell, Califano’s team used a novel algorithm to identify a “hidden” network of mRNAs and microRNAs that together control PTEN expression. They showed that 13 detectable genetic alteration of PTEN,” Califano says.

Bringing the Data to Biologists and Physicians

Experimental biologists and clinical researchers are in the best position to translate cancer genome findings into meaningful advances in patient care. But they often lack the expertise needed to access and make sense of the data. A team of scientists at Georgetown University is trying to bridge this gap by creating a user-friendly integrated database called G-DOC (Georgetown Database of Cancer), which they describe in a September 2011 paper in Neoplasia.

“We are unique in the way that we provide the data mining as well as the analytics in one environment,” says Subha Madhavan, MD, director of clinical research informatics at the Lombardi Comprehensive Cancer Center at Georgetown. G-DOC integrates patient data, genomic data, and small molecule data (“for matchmaking molecular targets with drugs,” Madhavan says) with popular tools for analyzing and visualizing these data, including GenePattern, Pathway Studio, and Cytoscape.

Continued on page 8
Users (both internal and external to Georgetown) have contributed clinical data on more than 3000 patients with breast or gastrointestinal cancers. “We were just lucky to work with investigators who had de-identified clinical data that we could leverage,” Madhavan says. Madhavan’s team has also imported a wealth of data from public databases and published articles. “We bring in the raw data and standardize them, so there’s a lot of value added to that data,” Madhavan says. They will add TCGA data for breast and colon cancer when they become available, she says.

Researchers can use G-DOC to generate or test hypotheses; run in silico experiments; learn about the newest types of data—including next generation sequencing, metabolomics, DNA copy number abnormalities, and microRNA expression—as well as about systems biology; and speed up the pace of their research. For example, it took one person using G-DOC one month to complete a colon cancer analysis that would otherwise have taken a team of people at least a year to complete, Madhavan says.

Overcoming Computing Barriers
To fully realize the vision of personalized cancer therapy, more labs will need to become computationally savvy, Califano concludes. “Right now there are a few computationally empowered labs, but the vast majority of other labs can only access computational analyses through collaborations,” Califano says. “There has to be some kind of connective tissue.”

“Right now ... the vast majority of labs can only access computational analyses through collaborations,” Califano says. “There has to be some kind of connective tissue.”

Gene expression signatures that stratify patients into likely and unlikely treatment responders are already in clinical use for certain cancers. But these “first generation” tests have severe limitations, says W. Fraser Symmans, MD, professor of pathology at MD Anderson Cancer Center. He and his colleagues are using state-of-the-art bioinformatics and biostatistics techniques to develop the next generation of gene expression tests for breast cancer.

Symmans and his colleagues discovered a paradox with some first generation tests for breast cancer. The tests accurately separate patients into “good” and “poor” responders to chemotherapy; but the “good responders” have worse survival. (The tests misclassify certain aggressive tumors that initially respond vigorously to chemotherapy but tend to relapse.) His team developed a second-generation test, described in the May 2011 issue of *JAMA*, that overcomes this issue and accurately predicts survival.

The test comprises a series of gene signatures (from the tumor) that sequentially predict: (1) response to hormonal treatment; (2) resistance to chemotherapy; and (3) sensitivity to chemotherapy. “We realized that one predictor was not going to be enough to capture the complexity,” says Christos Hatzis, PhD, who led the computational aspects of the project; Hatzis is founder and vice president of technology at Nuvera Biosciences Inc., which has commercial rights to the technology. The team used a multivariate approach to identify the key genes that define the signature; univariate approaches yield too many redundancies because genes work in pathways, Hatzis says.

The test accurately identifies patients who will respond to therapy about twice as often as standard methods. “It doesn’t completely solve the problem but it’s a big step forward,” Hatzis says.