SciBX
Science-Business eXchange

Gazing deeper into cancer

By Lev Osherovich, Senior Writer

Three recent genomic studies in glioblastoma multiforme,[1] acute myeloid leukemia[2] and lung adenocarcinoma[3] reveal potential new drug targets and provide a next step toward personalized, whole genome–based cancer diagnostics. Genomic sequencing could provide more robust data than existing cancer tests, which sample only a handful of the genes that could be mutated.

The next steps are to sequence multiple genomes, validate the markers and eventually build a unified theory of how mutational networks determine cancer risk, disease prognosis and response to treatment.

"The take-home message of this work is that, first and foremost, it can be done," because genomic sequencing costs have come down and speed has come up thanks to new technologies, said Elaine Mardis, associate professor of genetics and co-director of the Genome Center at **Washington University**. Mardis participated in all three studies and was the corresponding author on the AML study.

The studies were published in *Nature* and used **Illumina Inc.**'s high throughput sequencing platform, as well as technology from companies such as the **454 Life Sciences** subsidiary of **Roche**.

"We think these studies are a potential fundamental breakthrough in cancer, enabled by the low cost of sequencing human genomes," said Illumina CEO Jay Flatley. Illumina plans to sequence 50 more cancer genomes in 2009.

Sense of tumor

In the AML study, Mardis' team sequenced DNA from cancerous bone marrow cells as well as from healthy skin tissue obtained postmortem from a female AML patient who died from disease relapse.

The team then lined up the cancer cell sequences with corresponding DNA from the skin sample and the Watson[4] and Venter[5] genomes to find mutations in gene-coding regions relative to the control genomes.

The cancer cells primarily bore mutations in *FMS-like tyrosine kinase 3* (*FLT3*) and *nucleophosmin (nucleolar phosphoprotein B23, numatrin)* (*NPM1*; *B23*), two genes previously linked to AML. The study also identified eight more genes with a broad range of functions in cancer and metabolism, including two G protein–coupled receptors and surface proteins involved in adhesion and signal transduction.

On the basis of the single patient sequenced in the study, Mardis said it's hard to know which of these mutations drive AML pathogenesis. Whereas some of the genomic changes could contribute to disease, others could be irrelevant passengers that the mutation-prone cancer cells picked up along the way.

To sort the wheat from the chaff, Mardis is sequencing the complete genome of cancer cells from a second AML patient. She told *SciBX* that the second patient, who is still alive, also carries the FLT3 and NPM1 mutations. Comparing the sets of mutated genes between these AML patients could identify genes that influence whether or not the disease is fatal.

The findings and methods have not been patented, said Mardis.

Horizontal integration

In the lung cancer study, researchers sequenced 623 known cancer genes in tumor tissues from 188 patients and uncovered frequent mutations in several members of the ephrin family, including the EPH receptors EPHA3, EPHA5, EPHA7, EPHB1 and EPHB6. These receptor tyrosine kinase proteins had not previously been linked with lung cancer.

The most commonly mutated or amplified genes in this study included those encoding the epidermal growth factor receptors (EGFRs) HER3 (ERBB3) and HER4 (ERBB4).

EGFR is the target of cancer drugs Tarceva erlotinib from **OSI Pharmaceuticals Inc**, **Genentech Inc.** and Roche; Vectibix panitumumab from **Amgen Inc.** and **Takeda Pharmaceutical Co. Ltd.**; and Erbitux cetuximab from **Eli Lilly and Co.**'s ImClone Systems Inc. unit. Tarceva is marketed to treat non–small cell lung cancer (NSCLC) and pancreatic cancer and is in Phase I/II trials for GBM. Erbitux is marketed for colorectal cancer and squamous cell carcinoma of the head and neck (SCCHN) and is partnered with **Bristol-Meyers Squibb Co.** and **Merck KGaA**. Vectibix is marketed to treat colorectal cancer.

The lung cancer paper also found frequent mutations in known tumor suppressor genes such as *retinoblastoma 1* (*RB1*), *adenomatous polyposis coli* (*APC*), *ataxia telangiectasia mutated* (*ATM*) and *p53*. Other known oncogenes such as *K-Ras*, *fibroblast growth factor receptors* (*FGFRs*) and members of the *neurotrophic tyrosine kinase receptor* (*NTKR*) family were mutated.

The lung cancer study included Mardis' Washington University team together with researchers at the **Broad Institute** of **Massachusetts Institute of Technology** and **Harvard University** and collaborators from other institutions. It was led by Matthew Meyerson, associate professor of pathology at Harvard Medical School and an associate member of the Broad Institute, and Richard Wilson, professor of genetics and director of the Genome Sequencing Center at Washington University.

Vertical integration

The GBM study is the first report from a consortium called the Cancer Genome Atlas Research Network (TCGA), which was established in 2005 to collect and catalog genomic sequence, structure and functional data for every common tumor type (http://cancergenome.nih.gov/index.asp).

The consortium is a joint program of the **NIH**'s **National Cancer Institute** and the **National Human Genome Research Institute** (NHGRI).

The TCGA researchers applied a variety of genomic tools to paint an integrated picture of cancer-related genetic change. In addition to sequencing a subset of cancer candidate genes, the team examined 206

Box 1. Quantity and Quality.

Two new initiatives are tackling the challenges of teasing out disease-driving sequence variants from the natural genetic variation in humans and standardizing sequencing platforms to ensure reliability and reproducibility.

The 1,000 Genomes Project (http://www.1000genomes.org/page.php), whose goal is to obtain genomic sequences from around 1,000 individuals, was launched in 2008 by researchers at **The Wellcome Trust Sanger Institute**, the **Broad Institute** of **Massachusetts Institute of Technology** and **Harvard University** and other academic groups. It includes three sequencing companies: **Roche**'s **454 Life Sciences** subsidiary, **Illumina Inc.** and **Life Technologies Corp.**

In a pilot study, the consortium is making SNP maps and rough sequence drafts of the genomes of 186 people of diverse national origins. The genomes of about half a dozen individuals will be comprehensively sequenced to identify very rare mutations.

Richard Durbin, principal investigator at the Sanger Institute and cochair of the project, and collaborators at Illumina and various academic institutions published the first of these deep sequences—those of Yoruban[7] and Han Chinese[8] individuals—in *Nature* articles in November 2008 and released the complete high-density SNP maps of four individuals in December 2008.

Although the primary scientific goal of the project "is to help studies on the inherited genetics of common diseases," the resulting data will be useful for interpreting cancer genomic sequencing studies, said Durbin. "It's of interest to the cancer people for two reasons: to get background on normal variation and also because we're pushing the sequencing and data-handling methodology."

Meanwhile, Federico Goodsaid, associate director for operations in genomics at the **FDA**'s Center for Drug Evaluation and Research is leading the Sequencing Quality Control (SEQC) project.[9] That government, industry and academic consortium aims to develop methods for cross-checking sequence data.

SEQC is an outgrowth of Goodsaid's earlier MicroArray Quality Control (MAQC) project, which set cross-platform standards for microarray data.[10]

Groups using different genome-sequencing platforms need to be able to compare their results, said Goodsaid. "People are excited about next-generation sequencing but are confused about the different platforms," he said. "Companies are competing with each other to see which technologies become standardized."

According to Goodsaid, the rapid evolution of sequencing technologies makes it challenging to ensure that data are consistent between studies. The SEQC project hopes to preempt some of the confusion that initially beset microarray technology by establishing user-generated standards before next-generation sequencing technology becomes widely used. —*LZO*

GBM samples for changes in DNA copy number, methylation patterns and gene expression.

The group converged on mutations in *phosphoinositide 3-kinase* (*PIK3CG*; *PI3K*) homologs as a hallmark of GBM. The team found that a PI3K regulatory subunit called PIK3R1 bore mutations predicted to disrupt its interaction with the enzyme's inhibitory subunit, presumably leading to constitutive activation.

Researchers also correlated GBM-associated mutations with changes in gene expression and chromosomal instability, allowing cancer-related genes to be grouped into functional pathways. For example, tumors with activating mutations in EGFRs, mostly in *HER2* (*ERBB2*), had patterns of gene expression that were similar to tumors with deletions of downstream tumor suppressors such as *PTEN* (*MMAC1*; *TEP1*) and *neurofibromin 1* (*NF1*).

The GBM studies reveal "the extent to which the mutation spectrum peppers some of the oncogenic pathways that we already know about," said Joe Gray, director of the life sciences division at **Lawrence Berkeley National Laboratory** and one of the authors on the *Nature* GBM study. Instead of a single causal mutation, he said, "there's a variety of ways for cancer to activate these mutations."

Mardis said comparing the GBM and lung cancer sequences shows that although specific mutations vary between tumor types, most of the mutations affect growth-regulating and metabolism pathways, such as EGFR and p53, that are already implicated in many cancers.

"Those are all pathways that make sense," she said.

The study's results largely overlap with an earlier report in *Science* from a group led by researchers at **Johns Hopkins University**.[6] That study used high throughput sequencing technology from **Agencourt Bioscience Corp.** to examine mutations, gene copy number and transcription profiles in GBM tumors from 22 patients.

Although both studies identified many of the same genes as mutated or misregulated in GBM, the Johns Hopkins team identified a high rate of mutations in one gene, *isocitrate dehydrogenase* (*IDH1*), which was not flagged in the *Nature* study. Identifying the reasons for such discrepancies among large-scale, whole-genome studies to improve the reliability of these approaches for diagnostic and prognostic applications is already being tackled by an FDA-led sequencing standards consortium (*see* **Box 1, "Quantity and Quality"**).

Richard Durbin, principal investigator at **The Wellcome Trust Sanger Institute**, said the GBM sequencing studies illustrate a trade-off between precision and comprehensive coverage. The Johns Hopkins group sequenced well-studied oncogenes, whereas the Washington University team sequenced an entire genome, including regions that are uncharted territory save for rough sketches from the Human Genome Project and the Watson and Venter genomes.

"If you sequence the whole genome, as the Washington University group did, you find mutations all over the place, so it's hard to know which mutations are the important ones," said Durbin.

Thus, groups like the TCGA will "need to confirm that the things they find are functional," according to Durbin.

Functional integration

Cancer profiling is currently limited to testing for the presence of hereditary risk markers such as specific *breast cancer 1 early onset* (*BRCA1*) and *BRCA2* mutations or testing for selected tumor gene expression patterns to assess drug response and disease recurrence risk, such as *K-Ras* in EGFR-positive tumors. Integrating whole-genome sequence information with an assessment of changes in genome structure and gene expression over time could allow clinicians to predict initial cancer risk and to monitor the course of disease during therapy.

"In the future, you won't just profile tumors, but the individual as well," said Roland Wicki, director of strategic portfolio management for SOLiD at **Life Technologies Corp.** (formerly Invitrogen Corp.). "This will allow you to determine the therapy."

SOLiD, the company's integrated genomics platform, uses microarrays for both genomic sequencing and transcriptional profiling.

Wicki did caution that more cancer genomes must be sequenced before the mutations most relevant to disease progression can be picked out from normal genetic variants. It is not yet clear which of the mutations identified in the *Nature* and *Science* studies are predictive of either cancer risk or are relevant to disease progression.

Flatley said his company is addressing this issue by collaborating with academic groups to sequence 50 individual tumors along with patient-matched healthy tissue. Illumina's initial focus is on understanding the genetic changes that lead to cisplatin resistance in ovarian and gastric cancers.

Wicki also noted that the tumor genome is not the whole picture. Figuring out the significance of mutations in tumor development would also require monitoring transcription, alternative splicing, DNA methylation and other forms of integrative genomic analysis provided by the SOLiD platform.

"You cannot infer much just from the sequence," said Wicki. "You have to do functional studies."

Lawrence Berkeley's Gray agreed. "Until we assess how these mutations influence the pathophysiology, we may be fooling ourselves about which of these are important," he noted.

Gray also believes that integrative genomic analysis of tumors at various times during therapy could reveal the progressive accumulation of mutations that leads to drug resistance.

Revving the engines

Meanwhile, Illumina and its competitors are scaling up the speed and lowering the cost of comprehensive sequencing, with the goal of making it a diagnostic tool.

The main challenge will be to persuade clinicians, patients and insurers that genomic analysis can lead to better prediction of risk and therapeutic efficacy than single-gene tests or trial-and-error treatments. "We have to make the economics so compelling that it would appear to be one of the few ways for the healthcare system to save money," said Flatley.

Therapies such as Gleevec imatinib, marketed by **Novartis AG** to treat chronic myeloid leukemia (CML) and gastrointestinal stromal tumors (GIST), cost upwards of $42,000 a year.

Flatley believes the sweet spot for genomic sequencing as a clinical diagnostic will be $2,000–$3,000 per genome. The current cost of sequencing is about $100,000.

Sequencing the genomes of immune cells gathered from the blood ultimately could be a cheaper alternative to bone marrow biopsy, the current diagnostic procedure for many lymphatic tumors.

Washington's Mardis noted that her team's AML study cost $700,000 and nine months of work at its outset, but the second AML genome cost $200,000 and took only three months.

"For the third AML genome, we're looking at a cost of $100,000," she said. "This is a price point that the NHGRI has targeted for next-generation sequencing."

"We already have a $60,000 genome sequencing technology, and our $10,000 SOLiD 3 technology will be launched in early 2009," said Candia Brown, senior manager of strategic portfolio management for SOLiD at Life Technologies. "We're starting to work toward the $1,000 genome."

Osherovich, L. *SciBX* 2(1); doi:10.1038/scibx.2009.3
Published online Jan. 8, 2009

REFERENCES

1. The Cancer Genome Atlas (TCGA) Research Network. *Nature*; published online Sept. 4, 2008; doi:10.1038/nature07385
   Contact: Lynda Chin, Dana-Farber Cancer Institute, Boston, Mass.
   e-mail: lynda_chin@dfci.harvard.edu
2. Ley, T. *et al. Nature*; published online Nov. 6, 2008; doi:10.1038/nature07485
   Contact: Elaine Mardis, Washington University School of Medicine, St. Louis, Mo.
   e-mail: emardis@wustl.edu
3. Ding, L. *et al. Nature*; published online Oct. 22, 2008; doi:10.1038/nature07423
   Contact: Richard Wilson, Washington University School of Medicine, St. Louis, Mo.
   e-mail: rwilson@watson.wustl.edu
   Contact: Matthew Meyerson, Dana-Farber Cancer Institute, Boston, Mass.
   e-mail: matthew_meyerson@dfci.harvard.edu
4. Wheeler, D.A. *et al. Nature* **452**, 872–876 (2008)
5. Levy, S. *et al. PLoS Biol.* **5**, e254; published online Sept. 4, 2007; doi:10.1371/journal.pbio.0050254
6. Parsons, D.W. *et al. Science* **321**, 1807–1812 (2008)
7. Bentley, D.R. *et al. Nature* **456**, 53–59 (2008)
8. Wang, J. *et al. Nature* **456**, 60–65 (2008)
9. U.S. Food and Drug Administration. Notice of solicitation for the sequencing quality control project. <http://www.fda.gov/OHRMS/DOCKETS/98fr/FDA-2008-N-0644-n.pdf>
10. Canales, R.D. *et al. Nat. Biotechnol.* **24**, 1115–1122 (2006)

COMPANIES AND INSTITUTIONS MENTIONED

454 Life Sciences, Branford, Conn.
Agencourt Bioscience Corp., Beverly, Mass.
Amgen Inc. (NASDAQ:AMGN), Thousand Oaks, Calif.
Bristol-Myers Squibb Co. (NYSE:BMY), New York, N.Y.
Broad Institute, Cambridge, Mass.
Eli Lilly and Co. (NYSE:LLY), Indianapolis, Ind.
Food and Drug Administration, Rockville, Md.
Genentech Inc. (NYSE:DNA), South San Francisco, Calif.
Harvard University, Cambridge, Mass.
Illumina Inc. (NASDAQ:ILMN), San Diego, Calif.
Johns Hopkins University, Baltimore, Md.
Lawrence Berkeley National Laboratory, Berkeley, Calif.
Life Technologies Corp. (NASDAQ:LIFE), Carlsbad, Calif.
Massachusetts Institute of Technology, Cambridge, Mass.
Merck KGaA (Xetra:MRK), Darmstadt, Germany
National Cancer Institute, Bethesda, Md.
National Human Genome Research Institute, Bethesda, Md.
National Institutes of Health, Bethesda, Md.
Novartis AG (NYSE:NVS; SWX:NOVN), Basel, Switzerland
OSI Pharmaceuticals Inc. (NASDAQ:OSIP), Melville, N.Y.
Roche (SWX:ROG), Basel, Switzerland
Takeda Pharmaceutical Co. Ltd. (Tokyo:4502), Osaka, Japan
Washington University, St. Louis, Mo.
The Wellcome Trust Sanger Institute, Hinxton, Cambridge, U.K.