Institute reported in a January 2010 *Clinical Cancer Research* (CCR) study of 68 breast carcinomas. The other, an amplified metastasis gene, MTDH, or metadherin, predicts poor clinical outcome.

Researchers are also using CNAs to find new treatments. In the generally benign juvenile brain tumor pilocytic astrocytoma, CNAs led to mutated BRAF genes and new drugs that inhibit mitogen-activated protein kinase, an enzyme that BRAF regulates, Baudis said.

Likewise, an anaplastic lymphoma kinase (ALK) oncogene gain, or amplification, may render an inheritable form of neuroblastoma amenable to a U.S. Food and Drug Administration–approved lung cancer treatment, crizotinib. “We are always on the lookout for CNAs that may be treatment targets or prognostic indicators,” said Fish, who directs the Cohen Children’s Medical Center neuroblastoma program in New Hyde Park, N.Y. The 2008 discovery of ALK’s role in neuroblastoma was a “watershed moment,” he added, in the history of this difficult disease.

**Genetic Signatures**

Visible gains or losses—amplifications or deletions—in any of 22 pairs of human chromosomes form a visual CNA profile. Long and short chromosomal “arms” are labeled p and q, respectively; numbers define regions, or “bands,” containing specific genes; and decimals identify smaller regions called “sub-bands.” The prognostic indicators residing at 8q22–24 are, thus, genes in regions 22–24 on the short arm of chromosome 8. Researchers identified the genes by comparing that region in both tumor and normal cells to establish a baseline copy number profile.

Baseline profiles are important because “there’s no such thing as a completely normal individual,” said Xiaowu Gai, Ph.D., director of the biomedical informatics department at Loyola University Chicago Stritch School of Medicine. All chromosomes show copy number variations, most of which are perfectly normal. Gai and his team use specialized software to pick out the abnormal copy number variations by comparing pathological samples to baseline, or healthy, tissue samples, he explained.

Once established, the appropriately labeled normal and abnormal variations represent genetic fingerprints that bioinformatics programs such as arrayMap can display and analyze for a pathology report.

Sometimes CNAs offer the only fingerprint available. A pathologist analyzing a basal-like breast carcinoma, for instance, would not have the clinical assay signatures associated with HER2 and hormone receptor subtypes. But he or she would have a CNA fingerprint.

Gains at 6p12.3–23, 8q24.21–22, and 10p12.33–14 and losses at 4p15.31, 5q12.3–13.1, 5q33.1, 10q23.33, 12q13.13–3, 15q15.1, and 15q21.1 characterize this aggressive, poor-prognosis tumor, the Netherlands Cancer Institute team reported in their 2010 CCR paper.

For breast cancers with well-described genetic fingerprints such as the 70-gene prognosis signature, CNA profiles also serve as important supplements. Otherwise known as MammaPrint, the 70-gene prognosis signature can identify risks of metastasis and recurrence. CNA analysis improves MammaPrint’s accuracy, reported University of California, San Francisco, Helen Diller Family Comprehensive Cancer Center researcher Koei Chin, M.D., Ph.D., and coauthors in a 2006 *Cancer Cell* paper. After analyzing 272 primary breast tumors, they identified amplifications on chromosomes 8, 11, 17, and 20 associated with substantially worse prognosis.

The Netherlands Cancer Institute group refined Chin’s analysis in their CCR report. The 70-gene, poor-prognosis signature shows gains in chromosomal regions 3q26.33–27.1, 8q22.1–24.21, and 17q24.3–25.1, they found. Losses at 16q12.1–13 and 16q22.1–24.1 are associated with the 70-gene, good-prognosis signature.

“These correlations will help develop improved prognostic tests that may be used in patient-tailored therapy,” said study coauthor Marc van de Vijver, M.D., Ph.D., who directs the Netherlands Cancer Institute diagnostic oncology division.

**DNA Array**

To appeal to users without a bioinformatics background, Baudis and Institute for Molecular Life Sciences co-researchers Hyong Cheol Kim and Nitin Kumar built arrayMap “with a strong emphasis on a user-friendly data interface,” said Cai explained.

Indeed, arrayMap offers functionality that other genomics databases lack, said Yidong Chen, Ph.D., an epidemiology and biostatistics professor at the University of Texas Greehey Children’s Cancer Research Institute in San Antonio.

“Both GEO and ArrayExpress have limited visualization and analysis functions,” said Fishen. “The data visualization provided by arrayMap is a giant step forward.”

Visitors to http://www.arraymap.org find straightforward search tools, some requiring a username and password. “Search Samples” returns International Classification of Diseases for Oncology (ICD-O) codes, PubMed identification numbers, bioinformatics identifiers, and a text query function. A query on “breast carcinoma” returns several subtypes, including adenocarcinoma, infiltrating duct carcinoma, and inflammatory carcinoma. Clicking on “inflammatory carcinoma” returns a single genomic fingerprint from a 2006 study in *Cancer Cell*—a rare point of comparison for an inquiring pathologist. The array lays out gains and losses in all chromosomes from the inflammatory carcinoma cell lines used in the study. Click again on the long arm of chromosome 17, 17q, and the University of California, Santa Cruz, genome browser appears, showing that at least one well-known oncogene, the HER2-encoding ERBB2, may be part of inflammatory breast cancer’s genetic makeup.

The opportunity for integrated exploration beyond arrayMap’s virtual borders is one of its added values, Baudis explained. Like a librarian, it guides users through giant repositories that “adhere to formal standards but do not explore data content,” he said.

**Refining Treatments**

A “bizarre cancer” with a “spectrum of outcomes, from going away on its own to
fatality,” pediatric neuroblastoma is the most common extracranial solid tumor among infants, said Jonathan Fish, M.D., a pediatric oncologist.

Because of its highly individualized nature, neuroblastoma is also a prime candidate for CNA profiling, Fish explained. Neuroblastoma patients with aneuploidy in their tumors have a better prognosis. Oncogenic MYCN amplification on 17q indicates a worse prognosis. ALK amplification may offer both a diagnostic test and an entry to effective therapy.

Resources such as arrayMap, Fish said, “are valuable because they help refine treatments”—of particular importance among children, who have an 80% global cure rate. The child with a worse prognosis may benefit from more intense treatment; a brighter prognosis may mean a cancer that needs limited chemotherapy—or even disappears on its own.

“Two-thirds of children have a chronic illness that results from the cure—a complication of treatment,” explained Fish, who runs a long-term follow-up program for pediatric cancer survivors. “Refining treatment is especially important, particularly in children who may be living with its consequences for decades or have complications show up years later.”

**Universal Tool**

As an early step toward a universal genetic profiling tool, arrayMap offers one data store, Rossi said, that culls not only from public databases but also from trials and studies that might otherwise be obscured. Most clinical trials rely on patient-centered data that reads like “ductal breast carcinoma, stage I–II, premenopausal,” with little or no genomic information. arrayMap’s librarians continually screen the literature to match trial data with genetic arrays.

“At least half of all analyzed samples haven’t found their way into one of the big repositories like GEO or ArrayExpress,” Baudis explained. “We try to include this data in our literature collection, which currently has 65,450 cancer samples from 1,143 publications.”

Baudis is especially interested in rare but highly informative diseases. For example, “we are building a genomic data repository aimed at helping children with diffuse intrinsic pontine glioma, an uncommon but devastating childhood brain tumor,” he explained.

Though he believes that arrayMap is truly helpful, Rossi said it needs improvement, starting with the newest edition of the human genome (HG).

Otherwise known as “builds,” HG editions started in 2001 as HG-13. The latest edition, HG-19, reflects continual updating. arrayMap uses HG-18, Rossi said, which “can be problematic when you want all teams using the same, most-up-to-date build.”

Other experts would like different arrayMap refinements. The current version is hard to search and query, Chen explained. “The online software is not polished. It’s difficult to remember what each icon stands for, and while using the ICD-O code may be innovative for clinicians, researchers will be clueless what to do. We are too used to simple text search.”

Baudis partly agrees. “ICDO-3 is a good disease classification, but what are the best classification schemes?” he asked. “Can we develop a model of our own?”

An algorithm that excludes low-quality genomics data, which sometimes find their way into high-quality repositories, and “expert disease analysis to increase clinical usefulness” are also on Baudis’s list of “continuous improvements.” Meanwhile, he advises clinical oncologists to be ever mindful of oncogenomic data, which do not have proper priority in enough patient care scenarios.

“You can argue that CNAs and other genetic variations are actually underappreciated,” said Gai. “But they are related to so many diseases and syndromes that they’ve become absolutely important, especially as medicine becomes more personalized.”