Few in the medical science community would deny that, since the year 2000, concepts of disease and effective treatments have been greatly transformed by data that have given insight into the genomic basis of intracellular processes (giving strength to clinical goals of “personalized medicine”—now more accurately termed “precision medicine” [1]). The speed of that transformation is due in part to rapid advances in high-throughput DNA sequencing devices and the equally rapid dissemination of genome data by online journals and databases. Undoubtedly, a major accelerant has also been the dramatic decrease in the cost of whole-genome sequencing—from several billion dollars a dozen years ago to a mere few thousand dollars per genome now (2). This has even has spawned a direct-to-consumer genetic-analysis industry (eg, 23andMe [3]).

Nonetheless, gradually and in parallel with technical advances in gene sequencing, the early assumption that DNA sequencing alone could explain disease and help define specific therapies has given way to a more nuanced view that epigenetic processes, the cellular matrix, and aspects of the intracellular milieu such as micro RNA and proteomics are elements that modulate gene expression and are the scaffold of the diseased state (4). Many in the basic-science community admit, at least in a perfunctory way, that genotype manifests itself as phenotype (here read “imaging”) and that the one may inform the other. Also, there is increasing recognition that there is not only substantial heterogeneity between tumors but also within tumors (5). In one study (6), two-thirds of the mutations found in single biopsy samples were not uniformly detected in all the sampled regions of the same patient’s tumor. Here again, imaging has the potential to help non-invasively characterize the whole tumor, all tumors in the patient, and tumors at multiple time points over the course of treatment. However, there are few investigators who have yet pursued that potential connectivity in the published imaging literature. It is also true that federal program announcements and grant funding review committees have been slow to acknowledge or encourage investigation of that science intersection.

Despite the obvious complexity of disease biology, tests based exclusively on genetic data are proving to be clinically valued and are becoming U.S. Food and Drug Administration–approved products (eg, single pharmacogenetic markers for warfarin and statin sensitivity [7]). There are also gene cluster groupings for breast cancer such as OncotypeDx (8) (a 21-gene assay that predicts chemotherapy benefit and 10-year recurrence likelihood to guide treatment of women with early-stage invasive breast cancer) and Mammaprint (9) (a 70-gene signature that identifies which patients with early-stage breast cancer are at risk for recurrence after surgery). A failure of the imaging community to pursue research opportunities in this emerging field in a timely fashion would be unfortunate. Moreover, it might be pointed out that tissue “phenotype” features defined by using a careful lexicon and image analysis may already be discoverable through currently available advanced clinical imaging techniques (eg, clinical magnetic resonance [MR] imaging, dynamic contrast material–enhanced studies, fluorine 18 fluorodeoxyglucose positron emission tomography [PET]). Thus, initiating such research efforts may not need await development of future contrast agents or instruments.

The first and greatest hurdle for imaging researchers will be to achieve a grasp of the arcane language of genetic pathways and their therapeutically targetable pivot points (eg, p53,
EDITORIAL: Imaging and Genomics: Is There a Synergy?

Jaffe

mTOR, KRAS, VEGF, EGFR). A goal of future research engagement will be most productive if it focuses on populating a three-domain Venn diagram intersection made up of genetics, clinical data, and imaging features. To be statistically robust enough to be combinable with genetic data, researched imaging features need to be rigorously field tested for stability across multiple expert readers. And most importantly, the future of this research will need the inputs of network-connected interinstitutional cross-disciplinary teams of radiologists, computer image processing specialists, clinicians (eg, oncologists), statisticians familiar with integrated analysis of both imaging and genetic data types, and cellular physiologists.

So, in this new genetic universe, how can imaging develop a unique role and the resources needed to participate? Or will it be left behind? One promising route for imaging is exemplified by the article by Gevaert et al (10) in this issue. That study, which analyzed features extracted from non–small cell lung cancer (NSCLC) computed tomography and PET cases, offers an original approach to exploring the clinical prognostic value of imaging-genomics. Gevaert et al describe a 26-patient NSCLC cohort whose imaging features are comprehensively extracted (both observationally and objectively, by means of image processing), then statistically analyzed in light of their genomic properties. To address the need for the survival data that were not available to them from their own otherwise well-characterized cohort, they derived prognostic conclusions for their study by incorporating in their analytic model a genomically matched NSCLC case set with known clinical outcomes from public databases (11). This approach cleverly leverages both private and public data in a complementary way and advances the field by offering a promising new avenue for early exploratory imaging research that desires to study conventional clinical imaging relationships with cellular genomics. Moreover, the strength of their conclusions is made more convincing by their provision of extensive raw data in online supplements linked to their journal article.

A considerable assist in jump-starting investigator entry into this emerging field of science is offered by numerous National Institutes of Health (NIH)-sponsored data sharing resources. To reduce unnecessary experimental redundancy, the NIH has had a long-standing policy guidance that requests a data-sharing plan for recipients of large (>500,000) grants (12). Further growth of such resources will likely be enhanced by future grant review committees, whose policies will likely begin to include in their funding merit scores consideration of the applicant’s data-sharing plans. New multiagency federal initiatives encourage “Big Data” and their analysis as a means to speed up scientific discovery, especially at the intersection of complementary disciplines. Internet search engine sophisticated and community-minded information consolidators have made access to free, downloadable imaging analysis software (13) (often open source) no longer the barrier it once was. Self-forming interdisciplinary, dynamically communicating research teams have become further empowered by free or inexpensive online tools that permit document sharing, “Webinars,” meeting scheduling, and statistical analysis (eg, the R Project [14]). Nowhere is the intersection of all these online assets better illustrated than in Internet-connected teams formed to explore the heterogeneity of human cancer.

A major public data resource now presenting an important resource opportunity to imagers is an NIH program begun in 2003 that provides open access to high-quality genomic and clinical data on more than 20 tumor types. A data portal entitled “The Cancer Genome Atlas” (TCGA) (15) provides large-scale open access to genetic and proteomic data on 500 patient case collections for a range of more than 20 cancer tissue types (eg, glioblastoma, lung adenocarcinoma, ovarian, breast, prostate). In a leveraged effort intended to benefit imaging researchers, the National Cancer Institute Cancer Imaging Program (16) initiated a linked program entitled “The Cancer Imaging Archive” (TCIA) (17), which offers case-linkable diagnostic presurgical Digital Imaging and Communications in Medicine, or DICOM, images for a large subset of those same genomically analyzed cases. These images are coded by a common but private health information–protected identifier. Given the extensive and comprehensive genomic and clinical data (which in most cases include survival data and a great deal more) already residing in TCGA, image researchers might wish to perform ad hoc exploratory research that analyzes the images residing in TCIA with case-matched data in TCGA. Such a rich lode of freely accessible data resources and allied analytic engines (18,19) can readily serve as a starting point for any or all who wish to engage in this emerging science frontier. Initial efforts on TCIA-downloaded glioblastoma MR images joined with their TCGA-downloaded clinical and genomic data have already shown informative results (20). The fruitfulness of employing high-quality public data sets of this depth and breadth should be a substantial boon to researchers, because few individual institutions are able to assemble case cohorts of sufficient size to generate statistically robust results. Moreover, the scientific conclusions arising from such public data sets gain further scientific credibility because the open accessibility of their source data allows others to challenge or validate the conclusions they wish to do so.

Disclosures of Potential Conflicts of Interest: No potential conflicts of interest to disclose.

References

EDITORIAL: Imaging and Genomics: Is There a Synergy? Jaffe


