

The Cancer Genome Atlas Research Network: A Sight to Behold

Thomas J. Giordano

Published online: 1 November 2014
© Springer Science+Business Media New York 2014

Starting in 2010, I've enjoyed the distinct privilege to be the disease co-chair of the thyroid cancer project of The Cancer Genome Atlas (TCGA). This program is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). The program was conceived and proposed in 2005 (see http://cancergenome.nih.gov/PublishedContent/Files/pdfs/TCGA_executive_summary.pdf) and initiated in 2006 with several pilot projects, with the overarching goal being the comprehensive characterization of the somatic changes within the cancer genome. After initiation of pilot projects on brain, ovary, and lung cancers [1–3], TCGA was substantially expanded with a large fund infusion from The American Recovery and Reinvestment Act of 2009 (ARRA) to include the most common types of cancers including thyroid cancer. Here, my ultimate goals are to describe the key elements of the consortium and to discuss its relevance for the surgical pathology community, as we are in the early stages of a genomic transformation.

It was pure serendipity that I, with an introduction from Carolyn Compton, was first to knock on the TCGA thyroid door. Given my prior publications on gene expression profiles and genomic studies of thyroid cancer [4–6], TCGA leadership asked me if I would like to serve as the disease co-chair. I immediately accepted their offer, as TCGA represented the exact type of genomic science that wanted to pursue. My informatics co-chair was Gad Getz of the Broad Institute of Harvard and MIT and the Department of Pathology of Massachusetts General Hospital. It would become instantly obvious that Gaddy is a brilliant computational biologist, formally

trained in physics and then molecular biology. The project officially launched in May 2010 and we would share a wild journey over the next 4 years.

The first activity of TCGA projects is to form a Disease Working Group (DWG), a committee of disease experts whose primary mission is to determine the contents of the various data forms used to capture relevant clinicopathologic information. With input from TCGA leadership, we invited a group of accomplished endocrinologists, thyroid surgeons, and thyroid pathologists to join the DWG. Once assembled by May 2010, the DWG was also charged with defining the scope of the project, specifically deciding on which type(s) of thyroid cancer to study. TCGA strongly prefers to study one specific subtype of cancer in each project; for ovary they selected papillary serous adenocarcinoma [2]. The DWG initially proposed studying the full spectrum of follicular cell-derived thyroid cancers, from well-differentiated carcinomas (i.e., papillary and follicular) to poorly-differentiated and anaplastic/undifferentiated carcinomas. However, that idea was not acceptable to TCGA because this approach would likely dilute the statistical power needed to find new cancer genes. Looking back, there was a large degree of truth to that argument, but it would have also been beneficial to study in parallel a small cohort of histologically aggressive forms of thyroid cancers to permit direct comparisons between the different thyroid cancer types. Given that papillary thyroid carcinoma (PTC) is the most common type of thyroid cancer combined with the requirement for high-quality frozen tissues, selecting PTC for study was the only viable option.

The DWG of each project also guides, with assistance from the Biospecimen Core Resource (BCR, see below), the development of several forms to ensure standardized collection of pathology and clinical data. The *Initial Case Quality Control Form* captures initial pathology information. The *Enrollment Form* captures relevant clinical information and is only completed after the specimen passes key histologic and molecular

T. J. Giordano (✉)
Department of Pathology, Department of Internal Medicine,
University of Michigan Comprehensive Cancer Center, University of
Michigan Health System, 1150 West Medical Center Dr., MSRB1,
4520D, Ann Arbor, MI 48109-5602, USA
e-mail: giordano@umich.edu

metrics and is approved by the BCR for inclusion in the study. The *Follow-Up Form* captured outcome data and is only completed after 12 months has lapsed since case enrollment. These forms can be found at the BCR website (see <http://www.nationwidechildrens.org/biospecimen-core-resource-clinical-data-forms-standard>).

While the DWG was doing their work, the Biosample Core Resource (BCR) began receiving PTC samples and matched normal blood (non-neoplastic thyroid tissue from the contralateral lobe in cases with unifocal PTC was eventually permitted after it was determined to be essentially genetically indistinguishable from white blood cells) from Tissue Source Sites (TSSs), along with the data detailed in the Case Quality Control Form. Academic hospitals and commercial tissue banks served as TSSs, with IRB approval. The BCR for the thyroid project was located at The Research Institute at Nationwide Children's Hospital in Columbus Ohio under the leadership of Julie Gastier-Foster (see <http://www.nationwidechildrens.org/biospecimen-core-resource-for-the-cancer-genome-atlas>). The BCR was responsible for all aspects of tissue collection, storage, and handling, as well as nucleic acid extractions and distribution throughout the network for data generation. In addition, the BCR checks and verifies the clinical data before being uploaded to the Data Coordinating Center (DCC).

Cases were diagnosed as PTC by the pathologists from the respective submitting institutions. In addition, as part of the quality control process, the pathology was reviewed at the BCR with the assistance of several expert thyroid pathologists. Despite these efforts, the cohort is nonetheless still prone to the inter-observer diagnostic variability that is well documented to occur in thyroid cancer, especially the follicular variant of PTC [7–9].

The data generating capabilities of TCGA are extensive. Three Genome Sequencing Centers (GSCs) were established (Broad Institute, Baylor College of Medicine, and Washington University School of Medicine) to perform large-scale DNA sequencing using next-generation (NGS) sequencing technologies. Numerous Genome Characterization Centers (GCCs) were established to perform a variety of molecular assays to characterize the cancer genome. Each GCC had a special area of expertise, including copy number alterations (Broad and Harvard), epigenomics (Johns Hopkins and University of southern California), mRNA expression (University of North Carolina), miRNA analysis (British Columbia Cancer Agency), targeted sequencing (Baylor), and functional proteomics (MD Anderson Cancer Center).

The Data Coordinating Center (DCC) handled data management. In the genomic age in which genomes can be used to identify individuals [10], data security is vitally important; yet, one of the primary goals of TCGA is to make the data freely available to the cancer research community. Finding the correct balance between protecting patient's confidentiality and

providing access to the data was accomplished by developing policies in which different levels of data were made available with data use certification (see <http://cancergenome.nih.gov/abouttcga/policies/policiesguidelines>). TCGA also maintains a data portal for distribution of data (see <https://tcga-data.nci.nih.gov/tcga/>), as well as funds the Cancer Genomics Hub for distribution of primary sequencing data (see <https://cghub.ucsc.edu>).

After a significant number of PTCs completed the pipeline, the next step in the project was to form an Analysis Working Group (AWG). The AWG is responsible for executing the analysis and writing the first paper, referred to as the “marker” paper. Because the work of the AWG largely involves informatics and computational biology, the AWG consists mostly of analytical investigators together with a small group of disease experts from the DWG. Regrettably, this point represents the last involvement by most DWG members.

One of the main goals of the thyroid project was to identify cancer-initiating mutations, i.e., driver mutations, in those cases that lacked the well-known PTC driver mutations (*BRAF*^{V600E}, point mutations of *RAS* genes, and gene fusions involving *RET* and *NTRK*). These cases are referred to by computational biologists as “dark matter” cases, borrowing a term from physics. Because of the AWG's desire to explain as much of the dark matter as possible, we decided twice to wait to publish until all 500 cases went through the pipeline. It was largely these decisions that resulted in the project taking over 4 years to complete. It was a somewhat risky decision, as the publication embargo on the data expired well before submission of the manuscript. In the end, we made the correct decision to analyze and publish on a set of 496 PTCs because the larger cohort permitted several analyses that simply would not have been possible with a smaller cohort [11]. Furthermore, as predicted by TCGA leadership, the larger cohort allowed stronger statistical statements regarding the significance of some rare mutations. For example, we found 1.5 % of cases contained point mutations of *EIF1AX*. Using the MutSig algorithm [12], named for “mutation significance” and specifically designed to identify genes that are mutated more often than expected by chance alone given background mutation rates, we were able to conclude that these mutations were significant, and together with their mutual exclusivity with the common drivers, allowed us to suggest that *EIF1AX* mutation represents a novel driver mutation for PTC. We also found a diversity of gene fusions, such as novel *RET* partners and nine distinct types *BRAF* fusions in which *BRAF* undergoes a chromosomal rearrangement or translocation to create an oncogene. This study also recapitulated the established observations that (1) classical PTC is enriched for *BRAF*^{V600E} and *RET* rearrangements, (2) tall cell variant PTC is highly enriched for *BRAF*^{V600E}, and (3) follicular variant PTC is enriched for mutations of the *RAS* genes. It is interesting to note that, of the 99 follicular variant tumors, 19

tumors had some form of *BRAF* alteration; 12 had *BRAF*^{V600E}, and 7 had either small *BRAF* deletions, *BRAF* fusions, or *BRAF*^{K601E}. Our study clearly illuminated the various ways that *BRAF* can be mutated to become a PTC driver gene.

Along the way, the AWG faced two alternative paths regarding the scope of the project and marker paper. First, we could describe the essential findings of a few new cancer mutations and use the expression and DNA methylation data to report a molecular classification of PTC. However, the AWG wanted to leverage the vast available pan-genomic data and the outstanding computational talent of the AWG to tell a more complete, integrated, and biologically relevant story. So, we agreed to push the envelope. Given that the two main driver mutations of PTC, *BRAF*^{V600E} and *RAS*, were mutually exclusive allowed the derivation of a gene expression signature that could discriminate *BRAF*^{V600E} from *RAS*-mutant tumors. Using this signature, we derived a single measure, termed the *BRAF*^{V600E}-*RAS* score (BRS). The BRS was then used to interrogate the other mutations in the cohort to see if their biology resembled either *BRAF*^{V600E} or *RAS* mutant tumors. This analysis was very illuminating; for example, a single PTC with a *BRAF*^{K601E} was strongly *RAS*-like according to its BRS measure, suggesting that the *BRAF*^{V601E} protein kinase acts similar to a *RAS* mutation, termed *RAS*-like. Consistent with this is the observation that tumors with *BRAF*^{K601E}, as well as other *RAS*-like mutations, are highly enriched for follicular variant PTCs.

We then turned to thyroid differentiation because it plays a central role in thyroid cancer. We developed another gene expression-based measure termed the Thyroid Differentiation Score (TDS). Examining the relationship between TDS, the BRS and genotype allowed us to illustrate that *RAS*-like tumors were more differentiated than *BRAF*^{V600E}-like tumors. Furthermore, we showed that *BRAF*^{V600E}-like tumors showed a full range of TDS values, illustrating the heterogeneity of these tumors. This was highly significant because many studies have treated *BRAF*^{V600E}-mutant PTC as a homogeneous group.

We then used the BRS and TDS values to inform the molecular classification of PTC. Clustering of the gene expression (mRNA and miRNA), protein expression and DNA methylation data consistently showed a strong separation between *RAS*-like and *BRAF*^{V600E}-like tumors. The *RAS*-like group had high BRS and high TDS values, whereas the *BRAF*^{V600E}-like group showed the opposite pattern. Clustering of the *BRAF*^{V600E}-like group alone showed clusters with statistically distinct mean TDS values. Collectively, by using the BRS and TDS values in this manner, it was possible to argue that the resulting molecular classification of PTC, in which four subgroups of *BRAF*^{V600E}-like PTCs were identified, was biologically meaningful. Furthermore, using an integrated analysis of the miRNA data, we developed a hypothesis that high expression of miR-21 characterized the tall cell variant of PTC.

Our overarching conclusion was that *RAS*-like and *BRAF*^{V600E}-like PTCs have fundamentally different biology [11]. Since pathologic tumor classifications schemes should ultimately reflect biology, our result begs the question about whether these tumors belong together under the designation of PTC. Perhaps *RAS*-like tumors, which are characterized by the follicular variant, should no longer be considered a type of PTC. Alternatively, they could become their own diagnostic category or be merged with follicular carcinoma, where they were once classified in the 1960s before the follicular variant of PTC was recognized by Lindsay and then formally proposed by Chen and Rosai [13]. Regardless of which classification scheme is eventually derived, the TCGA thyroid cancer study clearly demonstrates how genomic information can illuminate the underlying biology of a tumor and impact its pathologic classification.

One of the goals of TCGA, in addition to characterizing the genome of the common cancers, was to construct a pipeline for team science through a geographically distributed network of diverse investigators with complementary skill sets. In this regard, the project is a resounding success. In doing this project, I personally learned so much and forged what I hope are lasting friendships. Besides managing a large and diverse group, I learned how to speak the basics of computational biology and spot a circular analysis. The project was certainly challenging, and it was difficult to write a linear manuscript of such a highly integrated analysis. Fortunately, I had adequate time to dedicate to this project, as I requested and was granted by Jay Hess and Jeff Myers a clinical sabbatical for the sole purpose of working on TCGA projects (I am also co-chair of the adrenocortical carcinoma project and a member of the pheochromocytoma/paraganglioma project). It was time well spent. Eventually, after many attempts, we found a way to structure the manuscript so that the story flowed from analysis to analysis [11].

The field of surgical pathology is clearly at a crossroads and programs like TCGA are leading the way. For example, the latest version of the sequencing assay for testing thyroid aspirates (ThyroSeq [14]) developed at the University of Pittsburgh has already incorporated findings from the TCGA thyroid project. While I am firmly in the camp that thinks that morphological assessment of tumors will remain the foundation of surgical pathology practice, by now it is unmistakably clear how molecular information can impact the classification of individual tumors and ultimately the treatment of patients. Surgical pathologists, in my opinion, must “own” the molecular evaluation of routine neoplastic surgical specimens and find ways to fully integrate the molecular and genomic data into our practices. I am confident that we can succeed. It would be ideal if we can eventually find a way forward in which the distinction between surgical pathology and molecular pathology becomes irrelevant.

Acknowledgments I thank my colleagues at the University of Michigan who covered much of my surgical pathology service while I pursued this project. I also thank TCGA leadership for the opportunity to serve as thyroid co-chair and Jean Claude ZenKlusen, TCGA Director, for reviewing the manuscript.

References

1. Cancer Genome Atlas Research N Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061–1068, 2008
2. Cancer Genome Atlas Research N Integrated genomic analyses of ovarian carcinoma. *Nature* 474: 609–615, 2011
3. Cancer Genome Atlas Research N Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489: 519–525, 2012.
4. Giordano TJ, Kuick R, Thomas DG et al. Molecular classification of papillary thyroid carcinoma: distinct BRAF, RAS, and RET/PTC mutation-specific gene expression profiles discovered by DNA microarray analysis. *Oncogene* 24: 6646–6656, 2005.
5. Giordano TJ, Au AY, Kuick R et al. Delineation, functional validation, and bioinformatic evaluation of gene expression in thyroid follicular carcinomas with the PAX8-PPARG translocation. *Clinical cancer research: an official journal of the American Association for Cancer Research* 12: 1983–1993, 2006.
6. Giordano TJ Genome-wide studies in thyroid neoplasia. *Endocrinology and metabolism clinics of North America* 37: 311–331, vii–viii, 2008.
7. Lloyd RV, Erickson LA, Casey MB et al. Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. *The American journal of surgical pathology* 28: 1336–1340, 2004.
8. Elsheikh TM, Asa SL, Chan JK et al. Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. *American journal of clinical pathology* 130: 736–744, 2008.
9. Duggal R, Rajwanshi A, Gupta N, Vasishta RK Interobserver variability amongst cytopathologists and histopathologists in the diagnosis of neoplastic follicular patterned lesions of thyroid. *Diagnostic cytopathology* 39: 235–241, 2011.
10. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y Identifying personal genomes by surname inference. *Science* 339: 321–324, 2013.
11. Cancer Genome Atlas Research N Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell* 159, 2014.
12. Lawrence MS, Stojanov P, Polak P et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499: 214–218, 2013.
13. Chen KT, Rosai J Follicular variant of thyroid papillary carcinoma: a clinicopathologic study of six cases. *The American journal of surgical pathology* 1: 123–130, 1977.
14. Nikiforov YE, Carty SE, Chiose SI et al. Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. *Cancer*, 2014.