The Pan-Cancer Atlas and TCGA’s Legacy

Group Meeting Talk 20/6/2018

Dvir Netanely
Ron Shamir’s Lab, Tel Aviv University
Welcome to the Pan-Cancer Atlas

From The Cancer Genome Atlas (TCGA) consortium, a large-scale collaboration initiated and supported by the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI).

From the analysis of over 11,000 tumors from 33 of the most prevalent forms of cancer, the Pan-Cancer Atlas provides a uniquely comprehensive, in-depth, and interconnected understanding of how, where, and why tumors arise in humans. As a singular and unified point of reference, the Pan-Cancer Atlas is an essential resource for the development of new treatments in the pursuit of precision medicine.
Covered publications:


The Cancer Genome Atlas: Creating Lasting Value beyond Its Data

• TCGA began in 2006 as a pilot project focused on three cancer types: lung, ovarian, and glioblastoma and was reauthorized for a full production phase in 2009.

• In the following decade, TCGA collected more than 11,000 cases across 33 tumor types and generated a vast, comprehensive dataset describing the molecular changes that occur in cancer.

• Tissue sample collection and data generation were completed in 2013 and 2016, respectively.

• Network marker papers, which are integrative cross-platform analyses of TCGA data on individual cancer types have been published for 31 of the tumor types to date.

The TCGA project in numbers

THE CANCER GENOME ATLAS (TCGA) BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data

To put this into perspective, 1 petabyte of data is equal to

212,000
DVDs

TCGA data describes

33
DIFFERENT TUMOR TYPES

...including

10
RARE CANCERS

...based on paired tumor and normal tissue sets collected from

11,000
PATIENTS

...using

7
DIFFERENT DATA TYPES

TCGA Results & Findings

RESULTS & FINDINGS

MOLECULAR BASIS OF CANCER
Improved our understanding of the genomic underpinnings of cancer

For example, a TCGA study found the basal-like subtype of breast cancer to be similar to the serous subtype of ovarian cancer on a molecular level, suggesting that despite arising from different tissues in the body, these subtypes may share a common path of development and respond to similar therapeutic strategies.

TUMOR SUBTYPES
Revolutionized how cancer is classified

TCGA revolutionized how cancer is classified by identifying tumor subtypes with distinct sets of genomic alterations.*

THERAPEUTIC TARGETS
Identified genomic characteristics of tumors that can be targeted with currently available therapies or used to help with drug development

TCGA’s identification of targetable genomic alterations in lung squamous cell carcinoma led to NCI’s Lung-MAP Trial, which will treat patients based on the specific genomic changes in their tumor.

THE TEAM

20 COLLABORATING INSTITUTIONS across the United States and Canada

WHAT’S NEXT?

The Genomic Data Commons (GDC) houses TCGA and other NCI-generated data sets for scientists to access from anywhere. The GDC also has many expanded capabilities that will allow researchers to answer more clinically relevant questions with increased ease.

*TCGA’s analysis of stomach cancer revealed that it is not a single disease, but a disease composed of four subtypes, including a new subtype characterized by infection with Epstein-Barr virus.
## SnapShot: TCGA-Analyzed Tumors

Amy Blum, Peggy Wang, and Jean C. Zenklusen
Center for Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Prevalence</th>
<th>TCGA cases assessed</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast lobular carcinoma</td>
<td>3,327,552</td>
<td>203</td>
<td>FOXA1 elevated in lobular carcinoma, GATA3 in ductal carcinoma; lobular enriched for PTEN loss and Akt activation</td>
</tr>
<tr>
<td>Breast ductal carcinoma</td>
<td>3,085,209</td>
<td>784</td>
<td>Four distinct subtypes: basal, Her2, luminal A, and luminal B; most common driver mutations: TP53, PIK3CA, GATA3; basal subtype similar to serous ovarian cancer</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>1,317,247</td>
<td>276</td>
<td>Highly heterogeneous with 26% driven by unknown alterations; ETS gene fusions or mutations in SPOR, FOXA1, or IDH1 define seven subtypes; actionable lesions in PIK3, MAPK, and DNA repair pathways</td>
</tr>
<tr>
<td>Colorectal adenocarcinoma</td>
<td>1,169,351</td>
<td>333</td>
<td>Colon and rectal cancers have similar genomic profiles; hypermutated subtype associated with favorable prognosis; new potential drivers: ARID1A, SOX9, Fam123B/WTX</td>
</tr>
<tr>
<td>Cutaneous melanoma</td>
<td>726,646</td>
<td>496</td>
<td>Established four subtypes: BRAF mutant, RAS mutant, NF1 mutant, and triple wild-type based on driver mutations; higher levels of immune lymphocyte infiltration correlated with better survival</td>
</tr>
<tr>
<td>Thyroid carcinoma</td>
<td>710,228</td>
<td>373</td>
<td>Majority driven by RAS or BRAFV600E mutations</td>
</tr>
<tr>
<td>Endometrial carcinoma</td>
<td>696,440</td>
<td>131</td>
<td>Classified endometrial cancers into four categories: POLE ultramutated, MSI (microsatellite instability) hypermutated, copy-number low, copy-number high</td>
</tr>
<tr>
<td>Uterine carcinosarcoma</td>
<td>527,228</td>
<td>230</td>
<td>Strong and varied degree of epithelial-mesenchymal transition; TP53 mutations in 91% of samples; PIK3 alterations in half</td>
</tr>
<tr>
<td>Invasive urothelial bladder cancer</td>
<td>527,228</td>
<td>178</td>
<td>Increased risk associated with smoking; frequently mutated; TP53 inactivated in 76% of tumors, ERBB2 (HER2), genes in the RTK/RAS pathways altered in 44%</td>
</tr>
<tr>
<td>Lung adenocarcinoma</td>
<td>483,225</td>
<td>446</td>
<td>High mutation burden; 76% have activation of receptor tyrosine kinase pathways</td>
</tr>
<tr>
<td>Lung squamous cell carcinoma</td>
<td>483,225</td>
<td>161</td>
<td>Commonly mutated genes: VHL, SED2, and the PIK3/AKT/mTOR pathway; metabolic shift similar to the “Warburg effect” correlates with a poor prognosis</td>
</tr>
<tr>
<td>Clear cell renal cell carcinoma</td>
<td>483,225</td>
<td>446</td>
<td>Extremely low mutation burden; metabolic shift distinct from the “Warburg effect” shift in clear cell carcinoma; TP53 and PTEN were frequently mutated; TERT gene promoter was frequently altered</td>
</tr>
<tr>
<td>Kidney papillary carcinoma</td>
<td>483,225</td>
<td>161</td>
<td>81% of type 1 tumors had MET alteration; type 2 tumors were heterogeneous, with alterations to CDKN2A, SETD2, TFE3, or increased expression of NRRF2-ARE pathway; loss of CDKN2A expression and CpG island methylation phenotype associated with poor outcome</td>
</tr>
<tr>
<td>Chromophobe renal cell carcinoma</td>
<td>483,225</td>
<td>66</td>
<td>Identification of HPV-negative, endometrial-like cancers with mutations in KRAS, ARID1A, and PTEN; amplification of CD274 and PDCD1LG2; frequent alterations in MED1, ERBB3, CASP8, HLA-A, and TGFBR2 and fusions involving IncRNA BCAR4; nearly three-quarters had alterations in either or both of the PI3K/MAPK and TGF-beta pathways</td>
</tr>
</tbody>
</table>

The Pan-Cancer Atlas: Motivation

- Cancers are typically classified using pathologic criteria that rely heavily on the tissue site of origin.
- However, large-scale genomics projects like TCGA produced detailed molecular characterizations of thousands of tumors, making a systematic molecular-based taxonomy of cancer possible.
- The analysis of molecular data have shown that each single-tissue cancer type can be further divided into several molecular subtypes.
The Pan-Cancer Atlas : Motivation

• To move toward a molecular taxonomy, the Pan-Cancer project investigated whether tissue-of-origin categories *split* or *converge* into pan-cancer subtypes based upon multiplatform genomic analyses.
  • What molecular alterations are shared across cancers arising from different tissues?
  • Do previously recognized disease subtypes in fact span multiple tissues of origin?
Combined analysis of 12 TCGA cancer types, 3,527 samples, 6 Omics.

Integrated 5-omic analysis using COCA identifies 11 big clusters: 5 were nearly identical to their tissue-of-origin counterparts, but several distinct cancer types were found to converge into common subtypes.

Lung squamous, head and neck, and a subset of bladder cancers coalesced into one subtype typified by TP53 alterations, TP63 amplifications, and high expression of immune and proliferation pathway genes.

Bladder cancers split into three pan-cancer subtypes.

Offers new classification – Better prognosis, shared drug targets.
Pan-Can 12 (2014) – The 11 COCA clusters
COCA (cluster of cluster assignment) (Hoadley et al., 2014)

• The algorithm takes as input the binary vectors that represent each of the platform-specific cluster groups and reclusters the samples according to those vectors.

• Advantages:
  • Data across platforms are combined without the need for normalization steps prior to clustering.
  • Each platform influences the final integrated result with weight proportional to the number of distinct subtypes reproducibly found by consensus clustering. Thus, “large” platforms (e.g., 450,000 DNA methylation probes) with orders of magnitude more features than “small” platforms (e.g., 131 RPPA antibodies) do not dominate the solution.
Pan-Can 12 (2014) – Prognostic value of the 11 COCA clusters

- COX analysis indicates that both tissue-of-origin and COCA subtype are both prognostic, and each provides independent information.
Breast cancers (BRCA) exhibit a pattern of divergence in which two main groups of samples are distinctly identifiable (Luminal + HER2 and Basal).

Whereas tissue of origin is the dominant signal for combined data on almost all of the other cancer types in the Pan-Cancer-12 collection, breast basal-like cancers are as different from luminal/ER+ breast cancers as they are from cancers of the lung.

The data from the present study strongly reinforce the idea that basal-like breast cancers constitute a unique disease entity.
Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer


Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer
Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer

• Previous Pan-Can analysis estimated that at least 10% of cancer patients might be classified (and perhaps treated) differently using a molecular pan-cancer taxonomy, rather than the current histopathology-based classification.

• Current questions:
  • Will the inclusion of many more tumors (x3) and tumor types enhance the number of cross-tissue associations?
  • Will it produce additional convergent and/or divergent integrated molecular subtypes?
  • Will it significantly increase the fraction of cancer patients whose classification or treatment might be affected by this new taxonomic approach?
PanCancer 33 (2018)

• This PanCancer study encompassed 11,286 tumor samples from 33 cancer types, for which molecular data were available from at least one of the five assay platforms.

• Of these, 9,759 had complete data for 4 platforms: aneuploidy (CNV), DNA methylation, mRNA and miRNA. RPPA protein data were available for a subset of samples (7,858).
Clustering by Individual Platforms

• Firstly, samples from each Omic were clustered separately using consensus hierarchical clustering.
• Resultant number of groups ranged from 10 to 25.
• While cell-of-origin was a dominant feature of the classification, the authors observed tumors from different cancer types grouping and samples within a cancer type dispersing across groups.
Samples were split mainly by those with few alterations (AN7), those with moderate alterations (AN6,8-10), and those with many alterations (AN1-5). Over one-third of the samples displayed relatively sparse aneuploidy in AN7; these were enriched for THCA, LAML, PRAD, and THYM. We observed more distinct clustering by cell-of-origin among higher-aneuploid tumors.
Despite the exclusion of loci known to be involved in tissue-specific DNA methylation, tumors originating from the same organ often aggregated by cancer-type-specific hypermethylation. This result suggests that cancer-associated DNA hypermethylation in human cancers is influenced by pre-existing cell-type-specific chromatin marks or transcriptional programs, and not just by cell-type-specific DNA methylation patterns.
Identified 25 groups that contained at least 40 samples. While tumor type was a driving feature for many groups, several groups were comprised of tumors from different organ types.
15 groups: While six groups contained only a single cancer type, the remaining nine groups each represented a mix of cancer types.
10 distinct protein groups: P1 (GBM, LGG) and P2 (DLBC, SARC, PCPG, UCS, THYM, and metastatic SKCM) were distinguished from the remaining 8 groups, largely corresponding to mesenchymal-like tumor types with high EMT signatures. Similar to the other individual data platforms, samples from related organ systems grouped together: luminal breast and gynecologic cancers (BRCA-Luminal, UCEC, and OV). In addition, a pan-kidney (P6) and a pan-GI (P8) group were identified.
Integrative Clustering across Data Types using COCA

- The clustering of cluster assignments (COCA) algorithm was used again to assess the overlap of platform-specific memberships from each of the five molecular platforms (aneuploidy, mRNA, miRNA, DNA methylation, and RPPA).
COCA- results

• Many samples grouped together by multiple platform-specific cluster memberships, both in groups that were defined by a single tumor type and in tumor types that co-clustered, such as KIRC and KIRP (pan-kidney).

• Gastrointestinal tumors (COAD, READ, STAD, and ESCA adenocarcinomas) co-clustered in the mRNA, miRNA, and RPPA platforms but were represented by several distinct DNA methylation clusters.

• Squamous histology cancers (LUSC, HNSC, CESC, ESCA, and BLCA) were similarly classified by the miRNA, mRNA and RPPA data but were further divided by the aneuploidy and DNA methylation data.

• Within pan-gyn cancers (BRCA, OV, UCEC, and UCS), RPPA data suggested that ovarian serous cystadenocarcinoma (OV) and UCEC (and ER+ LIHC) shared similarities at the protein level, whereas miRNA, mRNA, and DNA methylation data were grouped by their organ sites.

• Also of note, 13% of BRCA formed a subtype distinct from the majority of other BRCA, influenced by the mRNA and DNA methylation platforms.
Moving from COCA to iCluster

• While COCA showed high consistency across most data platforms, less concordance was observed for aneuploidy, where more than a third of the samples were defined by few to no aneuploidy events.

• COCA is less powerful when the molecular patterns are not strong enough to specify a distinct group on multiple individual platforms.

• To complement this analysis, joint clustering across all platforms simultaneously was explored.
iCluster analysis on PanCancer data

• We performed integrative molecular subtyping with iCluster using the four most complete data types (copy number, DNA methylation, mRNA, and miRNA) across 9,759 tumor samples, identifying 28 iClusters.

• The relative contribution of each platform to the overall clustering was quantified by summing the different platform feature weights on the iCluster latent variables.

• Copy-number alterations contributed 47% to the overall integrated clustering results, followed by the transcriptome (mRNA and miRNA) at 42%, and DNA methylation at 11%.
Unsupervised analysis using iCluster
Analyzing the resulting iClusters

iCluster robustness versus composition. Pie charts show the cancer-type composition within each iCluster and the size is proportional to the membership size. The cancer type accounting for the highest proportion of members within the iCluster was considered the dominant cancer type. The y coordinate of each pie center reflects this dominant cancer-type proportion; the x coordinate was determined by the iCluster silhouette width.
Analyzing the resulting iClusters

Relationship of TCGA tumor type, iCluster, and Pan-Organ system. The Sankey diagram demonstrates the tumor-type composition of each iCluster. The pan-cancer designations are shown on the right.
Tumor Maps of Organ Systems

• We visualized the samples by calculating Euclidean distances between the iCluster latent variables for all sample pairs and projecting the distances onto a 2D layout with TumorMap.

• More subtle differences within individual iClusters were apparent, potentially signifying important distinctions from the dominant cell-of-origin-associated signals.
Summary – PanCancer 33

• With nearly three times more tumors and tumor types profiled in this PanCancer Atlas analysis, the new analysis was able to detect more integrated molecular subtypes than reported in the original Pan-Cancer-12 analysis.

• While a third of iClusters were mostly homogeneous for a single tumor type, the other two-thirds showed varying degrees of heterogeneity.

• The most diverse group, C20:mixed (stromal/immune), contained a remarkable 25 tumor types. Most of the heterogeneous iClusters, including C20:mixed (stromal/immune), contained tumor types that fell within four major cell-of-origin, or organ system, patterns:
  • pan-GI
  • pan-gyn
  • pan-squamous
  • pan-kidney.
Summary – PanCancer 33

• Consequently, these four major cell-of-origin patterns are the subject of separate in-depth reports detailing their distinguishing genomic and molecular features.

• These iCluster assignments have potential clinical utility, and their multi-platform basis suggests that this new subclassification system might further improve the management of the 1%–3% of all cancer patients newly diagnosed with cancer of unknown primary (CUP).
A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers


https://www.cell.com/cancer-cell/fulltext/S1535-6108(18)30119-3
PanGyn - Motivation

• Gynecologic cancers share a variety of characteristics:
  • Arise from similar embryonic origins in the Müllerian ducts
  • Their development is influenced by female hormones
  • Managed by a particular medical specialty, gynecologic oncology

• Gynecologic and breast (Pan-Gyn) cancers have a projected incidence of more than 350,000 cases in the United States in 2017, with much larger numbers worldwide.

• Recently, similarities at the molecular level have been identified across gynecologic and breast cancers in a comprehensive analysis of all 33 TCGA tumor types

• Despite recent clinical advances, more comprehensive information on molecular characteristics of the tumors is a priority.

• The principal aims of the present study are to highlight both similarities and differences among types and subtypes of gynecologic cancers, in addition to the ways in which they differ from non-gynecologic cancers.

• Because breast tumors share most of the generic characteristics listed above, we have chosen to include them in the analysis.
PanGyn - Overview

• As part of The Cancer Genome Atlas (TCGA) Pan-Cancer Atlas project, we present here an integrated analysis of 2,579 patients' Pan-Gyn cancers at the DNA, RNA, protein, histopathological, and clinical levels.

• The study focuses on the following five TCGA tumor types:
  • high-grade serous ovarian cystadenocarcinoma (OV)
  • Uterine corpus endometrial carcinoma (UCEC)
  • Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC)
  • Uterine carcinosarcoma (UCS)
  • Invasive breast carcinoma (BRCA).

• We highlight shared characteristics and unique molecular features of the tumors, identifying clinically significant subtypes and suggesting potential therapeutic targets.

• Finally, we present a practical decision tree with only six laboratory-assessable molecular features, which classifies patient samples into one of five prognostic molecular subtypes.
PanGyn - Results

• Identified molecular features that distinguish Pan-Gyn from other tumor types:
  • 23 over mutated genes
  • 61 recurring SCNA (27 amp, 34 del) – some unique to PanGyn cancers
Mutations Distinguish Pan-Gyn from Other Tumor Types
SCNAs Distinguish Pan-Gyn from Other Tumor Types
The top five most frequently mutated genes were TP53 (44% of samples mutated), PIK3CA (32%), PTEN (20%), ARID1A (14%), and PIK3R1 (11%). Eleven of the 46 SMGs had not been previously reported in any of the TCGA gynecologic or breast marker papers.
Mutations Signatures in Pan-Gyn Tumor Types
Mutation Signatures

- Mutation signatures have provided insight into mechanisms underlying tumor development and have informed patient therapy (Helleday et al., 2014).
- Analysis by non-negative matrix factorization on the Pan-Gyn dataset suggested that 10 mutation signatures could explain nearly 90% of the variability observed in the original mutation/sample matrix.
Clustered Heatmap of Significantly Recurring SCNAs as Determined by GISTIC2.0 Analysis across Pan-Gyn Cancers
mRNA Expression Clusters and their Association with Overall Survival

(A) Unsupervised hierarchical clustering of previously reported cancer genes identifies nine mRNA-based subtypes/clusters. Clinical and molecular features are indicated by the annotation bars above the heatmap.

(B) Overall survival for each of the gene expression clusters (chi-square test p < 0.0001, adjusted for differences in tumor type survival rates).

(C) Overall survival for endometrial cancer (UCEC) patients in the gene expression clusters (log rank test p < 0.0001).

(D) Differential expression of ESR1, AR, SOX2, and CDH1 in different clusters (Kruskal-Wallis test p < 0.0001 for all four genes). The bars represent mean expression of the gene (log2 scale) in each cluster, together with the upper or lower 95% confidence interval (whiskers above or below the bars, respectively). See also Tables S4 and S5.

Figure 4. mRNA Expression Clusters and their Association with Overall Survival
COCA analysis (failed)

• We used cluster assignments from the six major TCGA platforms (mutations, SCNA, DNA methylation, mRNA, miRNA, and protein) to perform integrated clustering across the Pan-Gyn cohort using the CoCA.

• The resulting CoCA clusters were heavily dominated by tumor type because the intrinsic gene expression patterns were lineage dependent.

• Therefore, we turned to an alternative method to define subtypes that would span the Pan-Gyn tumor types and emphasize high-level similarities among them.
Subtypes across the Pan-Gyn Tumors

• We present molecular subtypes that illuminate commonalities and distinguishing features across the Pan-Gyn tumor types, with the potential to inform future cross-tumor-type therapies.

• We first identified 16 features across 1,956 samples that were either
  • (1) currently used in the clinic for at least 1 of the 5 tumor types, or
  • (2) identified as informative in previous TCGA gynecologic and breast cancer studies

• Next, we clustered the feature matrix and obtained 5 clusters.
Subtypes across the Pan-Gyn Tumors

Clustered heatmap of 16 features across 1,956 Pan-Gyn samples. Cluster 2 is split further into four subclusters, 2A–2D. Purple rectangles highlight HER2+ samples that have high immune infiltration scores; black rectangles highlight HER2+ samples with low immune infiltration scores.
Subtype Decision Tree

- Dichotomous decision tree methodology (Quinlan, 1983) was used to reduce the number of assessed molecular variables needed to classify patients into 1 of the 5 subtypes.
- The resulting tree required specification of only 6 of the original 16 features.
- The tree had an accuracy of 82% predicting the original 16-feature-based clusters, with a receiver-operator characteristic area under the curve of 0.94.
Cross-Tumor Type Pan-Gyn Subtypes with Prognostic Significance

(A) Clustered heatmap of 16 features across 1,956 Pan-Gyn samples. Cluster 2 is split further into four subclusters, 2A–2D. Purple rectangles highlight HER2+ samples that have high immune infiltration scores; black rectangles highlight HER2+ samples with low immune infiltration scores.

(B) Cross-tabulation showing the distribution of Pan-Gyn tumor types across the five clusters.

(C) Kaplan-Meier curves showing differences in overall survival among the five clusters (with 5- and 10-year survival rates shown). Before adjusting for tumor type differences in overall survival rates, the log rank test $p < 0.0001$, and after adjusting for tumor type differences, $p = 0.0006$ (chi-square test).

(D) Decision tree that predicts clusters using just 6 of the 16 features. The predicted clusters are shown in a covariate bar in the heatmap in (A).

(E) Kaplan-Meier curves showing differences in overall survival among the five decision tree-based predicted clusters (with 5- and 10-year survival rates shown). Log rank test $p < 0.0001$, before (log rank test) and after (chi-square test) adjusting for tumor type differences in overall survival rates. See also Figure S6; Tables S4 and S5.
Summary of main PanGyn findings

(1) Identified multiple genomic and epigenomic features that help to distinguish gynecologic and breast tumors from the other 28 TCGA tumor types;

(2) 61 somatic copy-number peaks in the Pan-Gyn cohort, 11 not previously reported by TCGA;

(3) 3 somatic copy-number alterations (containing genes of potential therapeutic relevance) unique to gynecologic cancers among the 33 TCGA tumor types;

(4) 46 SMGs in the Pan-Gyn cohort, 11 not previously reported by TCGA;

(5) 10 predominant mutation signatures, with 10% of the samples lacking identified SMGs;

(8) some OV and UCEC samples exhibited the “reactive” proteomic signature previously identified and shown to be prognostically relevant in BRCA;
Summary of main PanGyn findings

(9) identification of a subtype with low protein expression of ERs and AR (important markers for hormone therapy) that spanned all five tumor types;

(11) similar IncRNA profiles in BRCA and CESC, in contrast to the very distinct profiles in UCEC and OV;

(13) pathway analyses that revealed subsets of BRCA, OV, and UCEC samples with high levels of leukocyte infiltration, a primary marker of immune response and possible susceptibility to immunotherapy

(15) five cross-Pan-Gyn subtypes defined by multi-platform clustering of 16 molecular features; these five clusters have possible clinical implications and predictive value for survival beyond that of tumor type alone;

(16) reduction of the 16 molecular features to six in the form of a binary decision tree that retained prognostic value.
The TCGA Legacy

- The Cancer Genome Atlas (TCGA) represents a key milestone in NCI’s mission to reduce the burden of cancer suffering.
- Before TCGA, there was a growing sense in cancer research that we were nearing a complete understanding of the basic biology of cancer.
- TCGA ended that hubris, identifying whole new pathways that were very understudied in cancer biology.
- Examples: the role of the KEAP1-NRF2-CUL3 axis in squamous malignancies or the high frequency of SWI/SNF complex subunit mutations in many different types of cancer.

https://www.cell.com/cell/abstract/S0092-8674(18)30381-7
TCGA’s Legacy

• Improved the classification of multiple cancers
• Identified the molecular signature of each subtype.
• Revealed new drug targets
• Improved understanding of the various pathways underlying different cancer subtypes
• Motivated new algorithms and methods such as multi-omic analysis
• Teams are crucial for exploring such large datasets