

AWARD NUMBER: W81XWH-16-1-0237

TITLE: Analysis of Gastric Adenocarcinoma Data in a Pan-GI Context to Reveal Genes, Pathways, and Interactions that Yield Novel Therapeutic Advantages

PRINCIPAL INVESTIGATOR: Rehan Akbani

CONTRACTING ORGANIZATION: University of Texas MD Anderson Cancer Center
Houston TX 77030

REPORT DATE: September 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE September 2017			2. REPORT TYPE Annual		3. DATES COVERED 1 Sept 2016 - 31 Aug 2017	
4. TITLE AND SUBTITLE Analysis of Gastric Adenocarcinoma Data in a Pan-GI Context to Reveal Genes, Pathways, and Interactions that Yield Novel Therapeutic Advantages					5a. CONTRACT NUMBER	
					5b. GRANT NUMBER W81XWH-16-1-0237	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Rehan Akbani, Jaffer Ajani E-Mail: rakbani@mdanderson.org, jajani@mdanderson.org					5d. PROJECT NUMBER	
					5e. TASK NUMBER	
					5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Texas MD Anderson Cancer Center Houston, TX 77030					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012					10. SPONSOR/MONITOR'S ACRONYM(S)	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT The subject of this research is the study of gastric cancer, where the purpose is to reveal new insights into the biology of the disease that could potentially have therapeutic implications. Specifically, the scope of the study is based on 3 broad objectives: (i) identification of dysregulated and susceptible pathways, as well as their novel inter-relationships, in gastric adenocarcinoma (GAC); (ii) Pan-Cancer comparison of GAC with other cancers to leverage therapeutic target information across cancers; (iii) Identification of novel therapeutic targets, both with and without currently known drugs that target them. We have identified novel interactions amongst pathways in stomach and other cancers, where we have identified certain sub-groups of stomach cancer patients where those pathways may be exceptionally abnormal and lead to worse survival. The interactions and corresponding sets of genes may be targetable by existing drugs and/or drugs under development for treatment of that sub-group of stomach cancer patients. Other sub-groups have other interactions and genes that may also be targeted using different drugs. In that way, we can potentially give customized regimens of drugs to specific patients whose cancers exhibit targetable characteristics.						
15. SUBJECT TERMS Gastric cancer, disrupted pathways, targetable genes, PanCancer, Pan-gastrointestinal						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE	Unclassified	42	USAMRMC	
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER (include area code)	

TABLE OF CONTENTS

Introduction	4
Keywords	4
Accomplishments	4
Impact	17
Changes/problems	18
Products	18
Participants & other collaborating organizations	19
Appendices	21
Updated SOW	21
Manuscripts	24

INTRODUCTION

The subject of this research is the study of gastric cancer, where the purpose is to reveal new insights into the biology of the disease that could potentially have therapeutic implications. Specifically, the scope of the study is based on 3 broad objectives: (i) identification of dysregulated and susceptible pathways, as well as their novel inter-relationships, in gastric adenocarcinoma (GAC); (ii) Pan-Cancer comparison of GAC with other cancers to leverage therapeutic target information across cancers; (iii) Identification of novel therapeutic targets, both with and without currently known drugs that target them. We have identified novel interactions amongst pathways in stomach and other cancers, where we have identified certain sub-groups of stomach cancer patients where those pathways may be exceptionally abnormal and lead to worse survival. The interactions and corresponding sets of genes may be targetable by existing drugs and/or drugs under development for treatment of that sub-group of stomach cancer patients. Other sub-groups have other interactions and genes that may also be targeted using different drugs. In that way, we can potentially give customized regimens of drugs to specific patients whose cancers exhibit targetable characteristics. The objective of this research is to find those targets and improve our understanding of the biology of stomach cancer.

KEYWORDS

Gastric cancer, stomach cancer, gastrointestinal cancers, pathway aberrations, gastric cancer therapeutic targets, Pan-Cancer, dysregulated pathways, gastric cancer subtypes.

ACCOMPLISHMENTS

What were the major goals of the project?

The major goals of the project and their breakdown into milestones (as stated in the approved SOW) are shown below. Also shown is the percentage of completion for the milestones, to date. Please note that the milestones have not been completed in the originally proposed chronological order and Specific Aims 2 & 3 were worked on before Specific Aim 1. That was due to external factors beyond the PI's control, as described under the section "CHANGES/PROBLEMS." However, even though the order of completion has changed, the tasks themselves and the amount of time needed to complete them have not changed.

Specific Aim 1	Timeline	Percentage Completed
Major Task 1: Acquisition and Quality Control of gastric cancer data, in preparation for computational analysis	Months	
Acquire TCGA gastric cancer data, and in-house MD Anderson data (after procuring necessary approvals)	0.5	
Convert all acquired data into a "standardized" format suitable for computational analysis	1	
Assess and remove (if needed) batch effects from within TCGA and within MD Anderson data, and	2	

improve the overall quality of each data set individually.		
Merge TCGA data with MD Anderson data, removing batch effects across both data sets.	2.5	
Re-assess the quality of the overall data and iterate back to previous steps, if needed, until data are satisfactory	3	
Milestone(s) Achieved: “Cleaned up” gastric cancer data from TCGA and MD Anderson ready for computational analysis	3	
Major Task 2: Computational analysis of the gastric cancer data sets		
Cluster the data sets and study the results	4	
Generate pathway activity scores for various pathways across multiple data types, and determine which ones are likely disrupted	5-6	
Correlate disrupted pathways across multiple data types (e.g. transcriptomic, proteomic, genomic, epigenomic) and across clinical variables (e.g. histology, stage, grade, outcome) via statistical analysis	7-8	
Milestone(s) Achieved: First round computational analysis for gastric cancer completed	8	
Major Task 3: Publish gastric cancer results		
Discuss results with collaborators and perform any follow up analysis	9-10	
Write one or more manuscript(s) with input from designated mentor and collaborators	11-12	
Submit manuscript(s) and wait for reviews.	13-14	
Respond to reviewers and resubmit. May repeat submission/resubmission process with multiple journals depending on where the paper(s) end up being published. Present results at conferences.	15-19	
Milestone(s) Achieved: Manuscript(s) published	19	
Total time for Specific Aim 1	19	

Specific Aim 2		
Major Task 4: Acquisition and Quality Control of Pan-GI data, in preparation for computational analysis		
Acquire TCGA Pan-GI data and convert them into a “standardized” format suitable for computational analysis	20	100%
Assess and remove (if needed) batch effects from the data, and improve the quality of the data	21	100%
Milestone(s) Achieved: “Cleaned up” TCGA Pan-GI data ready for computational analysis	21	100%
Major Task 5: Computational analysis of the Pan-GI data sets		
Cluster the data sets and study the results	22	100%
Generate pathway activity scores for various pathways across multiple data types, and determine which ones are likely disrupted	23-24	80%
Correlate disrupted pathways across multiple data types (e.g. transcriptomic, proteomic, genomic, epigenomic) and across clinical variables (e.g. histology, stage, grade, outcome) via statistical analysis	25-26	80%
Compare gastric with other Pan-GI cancers and look for similarities and differences	27	100%
Milestone(s) Achieved: First round computational analysis for Pan-GI cancers completed	27	90%
Major Task 6: Publish Pan-GI cancer results		
Discuss results with collaborators and perform any follow up analysis	28-29	100%
Write one or more manuscripts with input from designated mentor and collaborators	30-31	100%
Submit manuscript(s) and wait for reviews	32-33	100%
Respond to reviewers and resubmit. May repeat submission/resubmission process with multiple journals depending on where the paper(s) end up being published. Present results at conferences.	34-36	50%

Milestone(s) Achieved: Pan-GI manuscript(s) published	36	87.5%
Total time for Specific Aim 2	36	
Specific Aim 3		
Major Task 7: Identification and publication of potential therapeutic targets in gastric cancer		
Identify potential genes and/or pathways in gastric cancer for targeted therapy, using gastric data only from Aim 1	9-10	
Identify potential genes and/or pathways in gastric cancer for targeted therapy, using cross-tumor information from Pan-GI cancers from Aim 2	28-29	100%
Integrate results into manuscripts for Specific Aims 1 and 2. Present results at conferences.	30-36	50%
Milestone(s) Achieved: Potential therapeutic targets identified and published	36	50%
Total time for Specific Aim 3 (interspersed with other aims; not consecutive months)	36	

What was accomplished under these goals?

Specific Aim 2, Major Task 4: Acquisition and Quality Control of Pan-GI data, in preparation for computational analysis

That aim has been fully completed and the following milestone has been achieved: “Cleaned up” TCGA Pan-GI data ready for computational analysis. In fact, not only have the Pan-GI data been adjusted for batch effects and standardized, but data across all 33 TCGA tumor types have been adjusted. Specifically, Dr. Akbani’s lab adjusted mRNA, miRNA and protein data, and the adjusted datasets are being used by the TCGA PanCanAtlas project as the “official” cleaned up datasets. The adjusted data are currently available at the password protected Synapse.org page for PanCanAtlas, but they will be released to the public at the Genomic Data Commons portal (gdc.cancer.gov) once the PanCanAtlas papers have been published in Spring 2018.

The following figures illustrate examples of batch effects that were found in the TCGA mRNA and miRNA data, as well as the results after correction by Dr. Akbani’s lab. (Key: gastric (STAD), colon (COAD), rectal (READ), esophageal (ESCA), AML (LAML), endometrial (UCEC) cancers, Stratagene reference (Strat)).

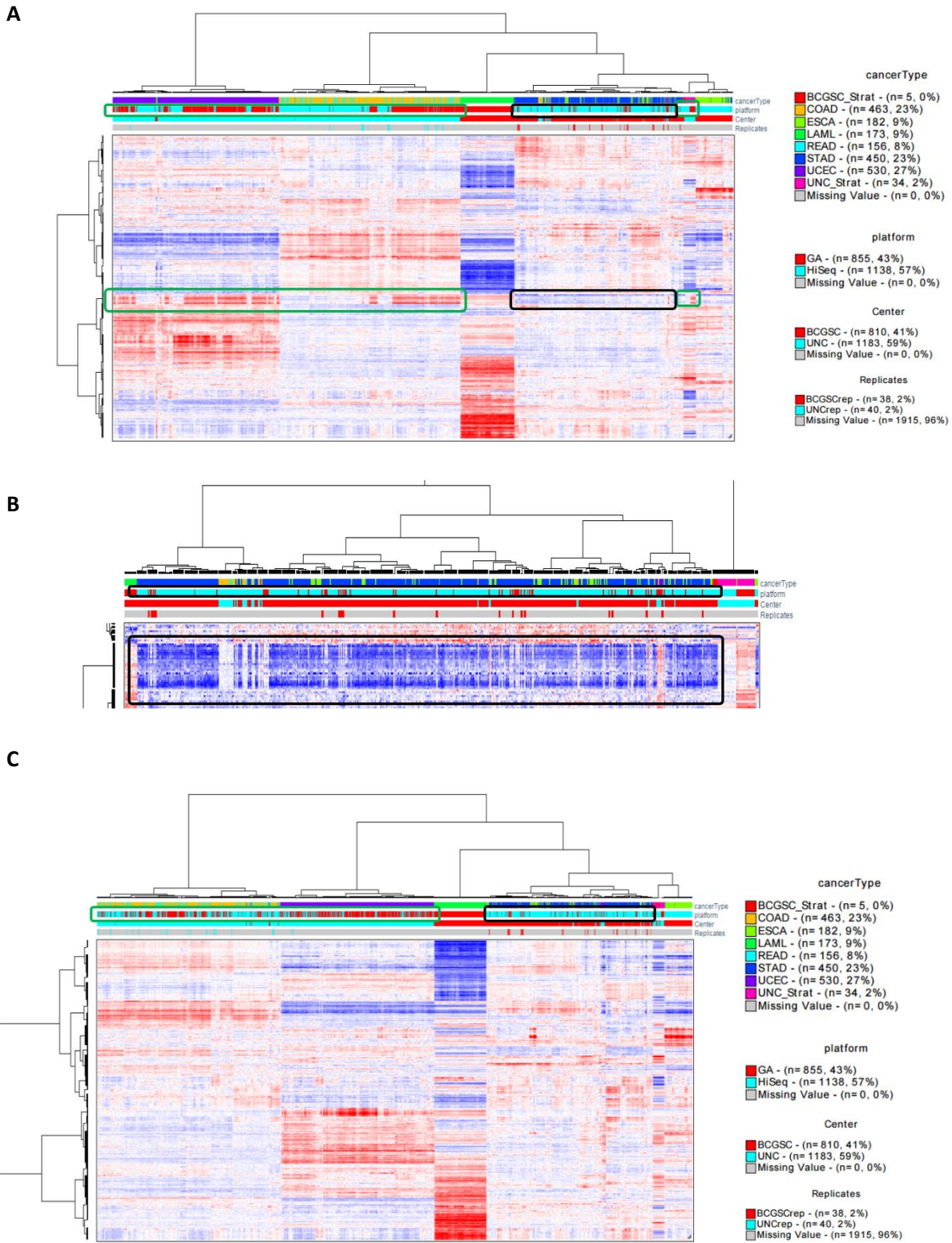


Fig. 1. (A) Clustered heat map of mRNA data with genes in rows and samples in columns (red = high, white = medium, blue = low expression). Large batch effects by platform can be seen in the green rectangles in COAD, READ, UCEC and Strat data. Subtler batch effects by platform are observed in gastric (STAD) data

(black rectangles). **(B)** Zoomed in view of the black rectangle in A, showing expression of certain genes corresponds with platform variable, demonstrating a batch effect. **(C)** Clustered heat map of the batch effects adjusted data. Batch effects by platform have been mitigated and the platform types can be seen to merge together (green and black rectangles).

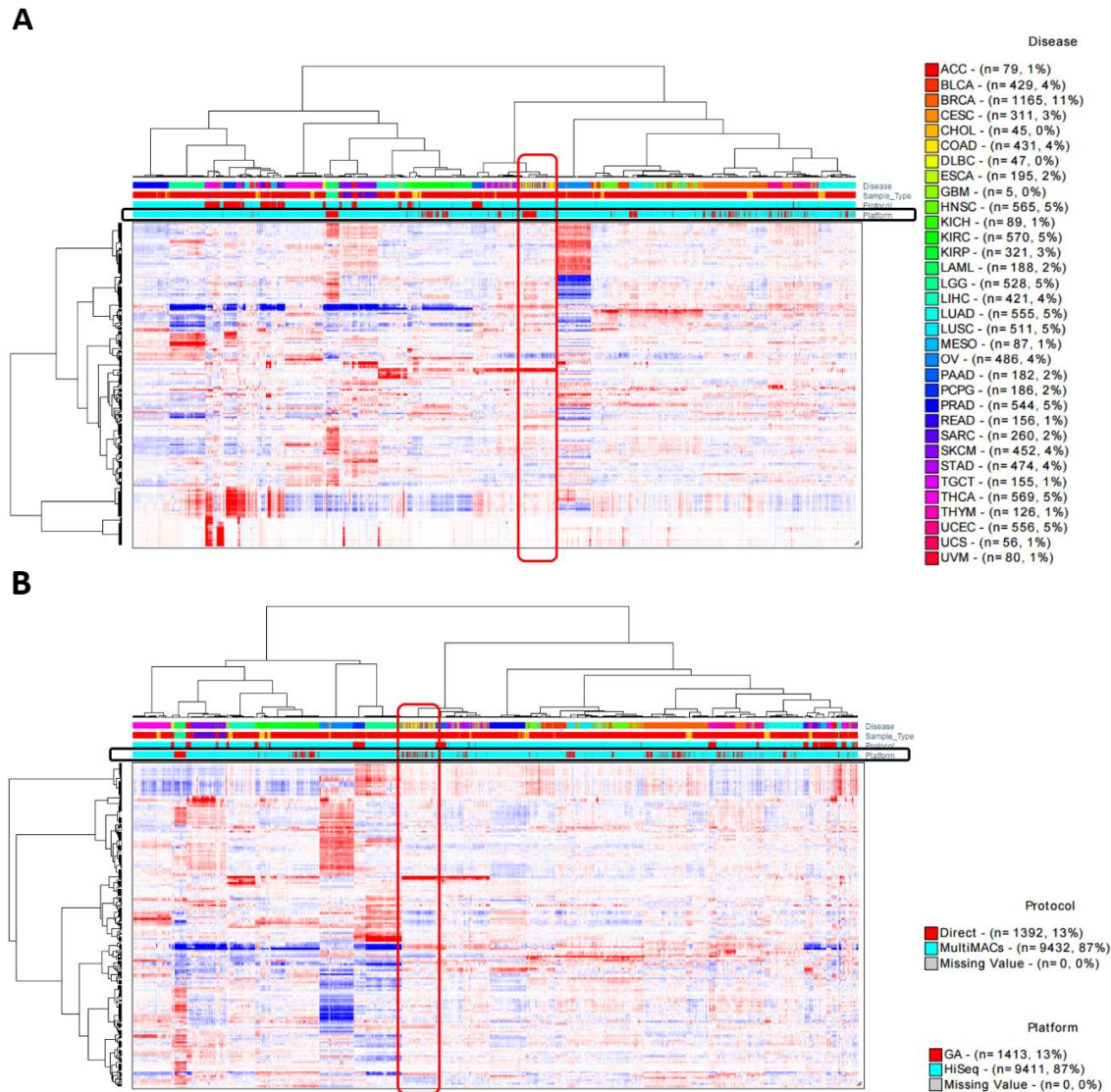


Fig. 2. (A) Clustered heat map of miRNA data with miRNAs in rows and samples in columns (red = high, white = medium, blue = low expression). Batch effects by platform can be seen (black rectangle), which is strongly illustrated in the colorectal data (red rectangle) where data from the two platforms do not merge. **(B)** Clustered heat map of the batch effects adjusted miRNA data. Batch effects by platform have been mitigated and the platform types can be seen to merge together (black and red rectangles).

Specific Aim 2, Major Tasks 5, 6 and 7: “Computational analysis of the Pan-GI data sets,” “Publish Pan-GI cancer results,” and “Identification and publication of potential therapeutic targets in gastric cancer.”

Most of those tasks have been completed. One manuscript has been published in Cancer Cell, another is currently under review by Cell and Cancer Cell (first round of reviews have been received and they look encouraging), whereas a third manuscript is currently being prepared and scheduled to be submitted to Cancer Cell in early November (see Appendices for references). Dr. Akbani is the corresponding author on the last one. Highlights of the major gastric cancer related findings in each of the manuscripts, respectively, are:

1) *A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations (Cancer Cell)*

- i. PI3K/AKT/mTOR pathway is disrupted in gastric cancer (Fig. 3).
- ii. *PIK3CA* is the most mutated gene in that pathway, with approximately 20% of the gastric samples having mutations in it (Fig. 3).
- iii. *DEPTOR*, *PIK3CA* and *RICTOR* are the most frequently amplified genes with approximately 5% of the gastric samples having amplifications in them (Fig. 3).
- iv. *PTEN* is the most frequently deleted gene in that pathway, with approximately 5% of the gastric samples having deletions in them (Fig. 3).
- v. Gastric cancer (STAD) has intermediate level of activities of the PI3K/AKT and mTOR pathways compared to other cancers (Fig. 4B).
- vi. It has high levels of phospho-mTOR, phospho-S6, and phospho-4EBP1 proteins, hinting at potential targets for therapy (Fig. 4A).

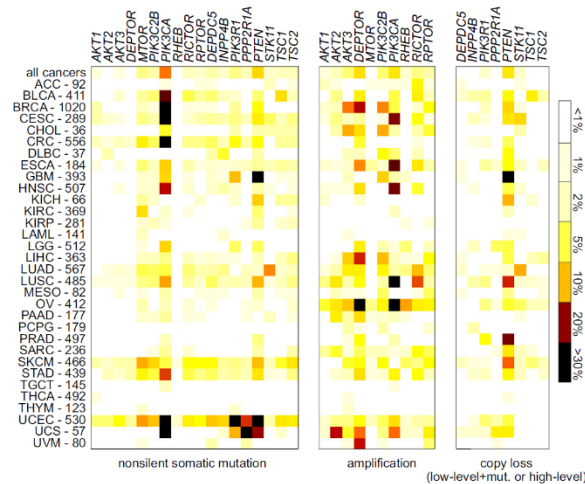


Fig. 3. (Adapted from Fig. 2B in the paper.) By cancer type, percentages of somatic mutation or copy alteration for each indicated gene. Amplification denotes “high-level” copy gain. Numbers of cases denote representation on Whole Exome Sequencing data platform.

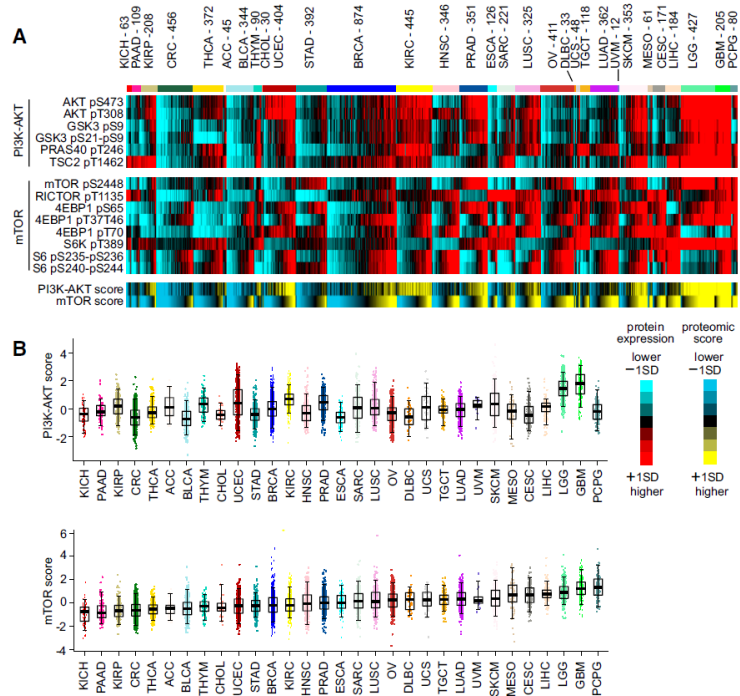


Fig. 4. (Adapted from Fig. 1A-B in the paper). Proteomic Signatures of PI3K/AKT and mTOR across Human Cancers **(A)** Heatmap of RPPA features considered core to either PI3K/AKT or mTOR pathways across 7,663 cancers. Red, higher expression (values normalized to SDs from the median across all cancers); blue, lower expression. PI3K/AKT and mTOR features were each summarized into pathway activity scores for each tumor profile (yellow, higher inferred activity; blue, lower activity; bright yellow/blue denotes change of 1 SD or SD, from the median). Cancer types (denoted by TCGA project name) are ordered by low to high average mTOR pathway score. **(B)** Boxplots of PI3K/AKT (top) and mTOR (bottom) pathway activities scores, as inferred using RPPA data. Boxplots represent 5%, 25%, 50%, 75%, and 95%.

2) Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas (under review by Cell and Cancer Cell)

- i) FBXW7, SMAD2, SOX9, MUC6 and ZFP36L2 are some of the genes that are significantly mutated in GI cancers, but not in non-GI cancers (Fig. 5A).
- ii) KRAS, GATA6, CDK6 and GATA4 are some of the genes that are significantly amplified in GI cancers, but not in non-GI cancers (Fig. 5B).
- iii) GMDS, PARK2, and SMAD4 are some of the genes that are significantly deleted in GI cancers, but not in non-GI cancers (Fig. 5B).
- iv) Proposed 5 novel Pan-GI subtypes; EBV+, hypermutated SNV, hypermutated indel, chromosomal instability (CIN), genomically stable (GS) (Fig. 6A).
- v) EBV+ is only found in gastric cancer. It also has a smaller proportion of CIN subtype compared to other GI cancers (Fig. 6B-C).
- vi) Gastric cancer has hypomethylation and fewer CpG island methylator phenotype (CIMP) samples than other GI cancers (Fig. 6C).
- vii) The upper GI tract has more AA > AC mutations than the lower GI tract (Fig. 6C).
- viii) RTK/RAS/PI3K, TP53/Cell cycle, TGF-beta, and WNT pathways are all genomically altered in GI cancers, providing potential avenues for targeted therapy (Fig. 7).

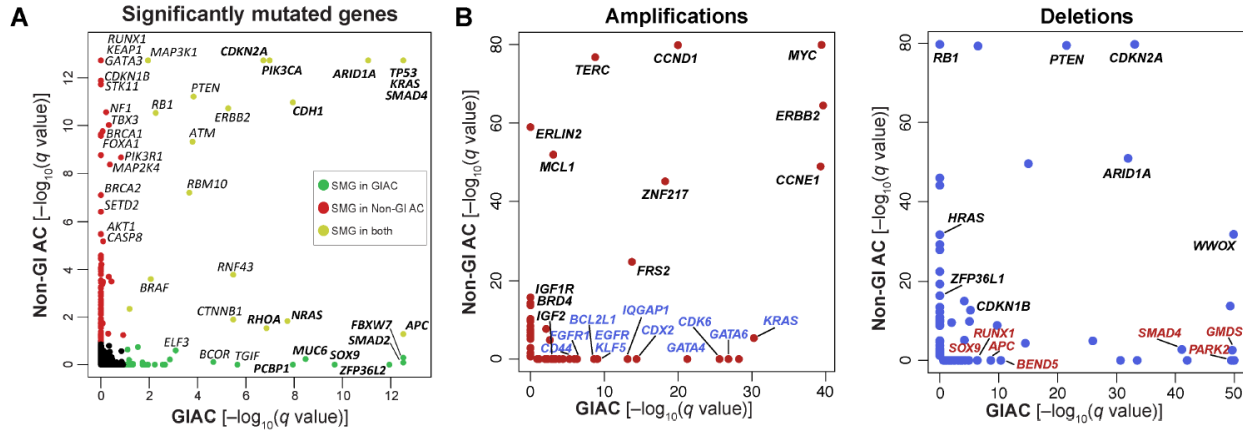


Fig. 5. Genomic features of gastrointestinal adenocarcinomas. **(A)** Quantile-quantile (Q-Q) plot of significantly mutated genes in gastrointestinal adenocarcinomas (GIAC) (horizontal axis) compared to other adenocarcinomas (non-GI AC) (vertical axis). Significantly mutated genes unique to GIAC are marked green flanking the horizontal axis, those unique to other adenocarcinomas are marked red bordering the vertical axis, and common to both are marked yellow scattered along the diagonal. **(B)** Q-Q plot of significant focal amplifications (left) and deletions (right) of GIAC compared to non-GI AC.

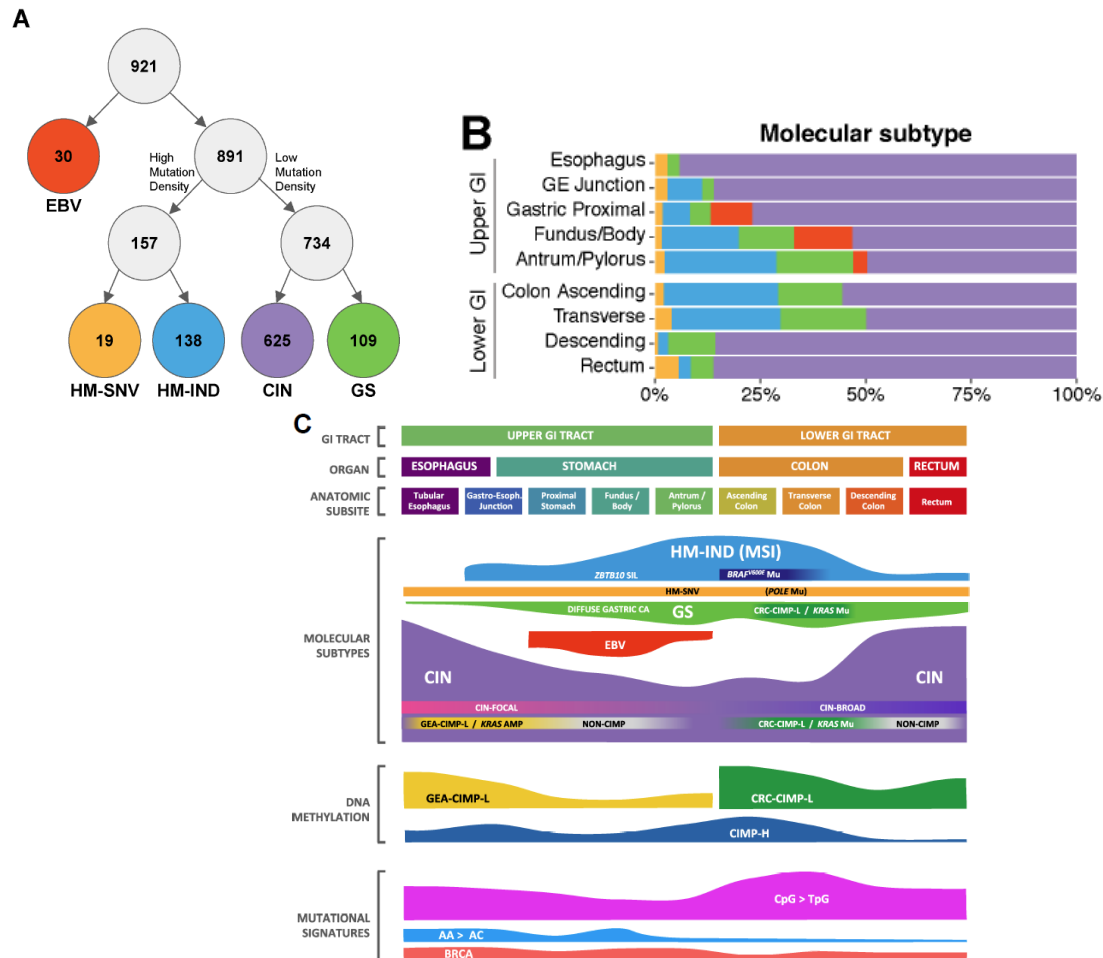


Fig. 6. Molecular subtypes of gastrointestinal adenocarcinomas. **(A)** Flowchart depicting the classification of GIAC molecular subtypes: Epstein-Barr virus (EBV)-positive (red); hypermutated-single-nucleotide variant predominant (HM-SNV) (gold); hypermutated-insertion/deletion predominant (HM-IND) (blue); chromosomal instability (CIN) (purple); and genomically stable (GS) (green). **(B)** molecular subtypes among GI cancers across the gastrointestinal tract represented by percentage per anatomic region. **(C)** Schematic summarizing key molecular features of GI cancers.

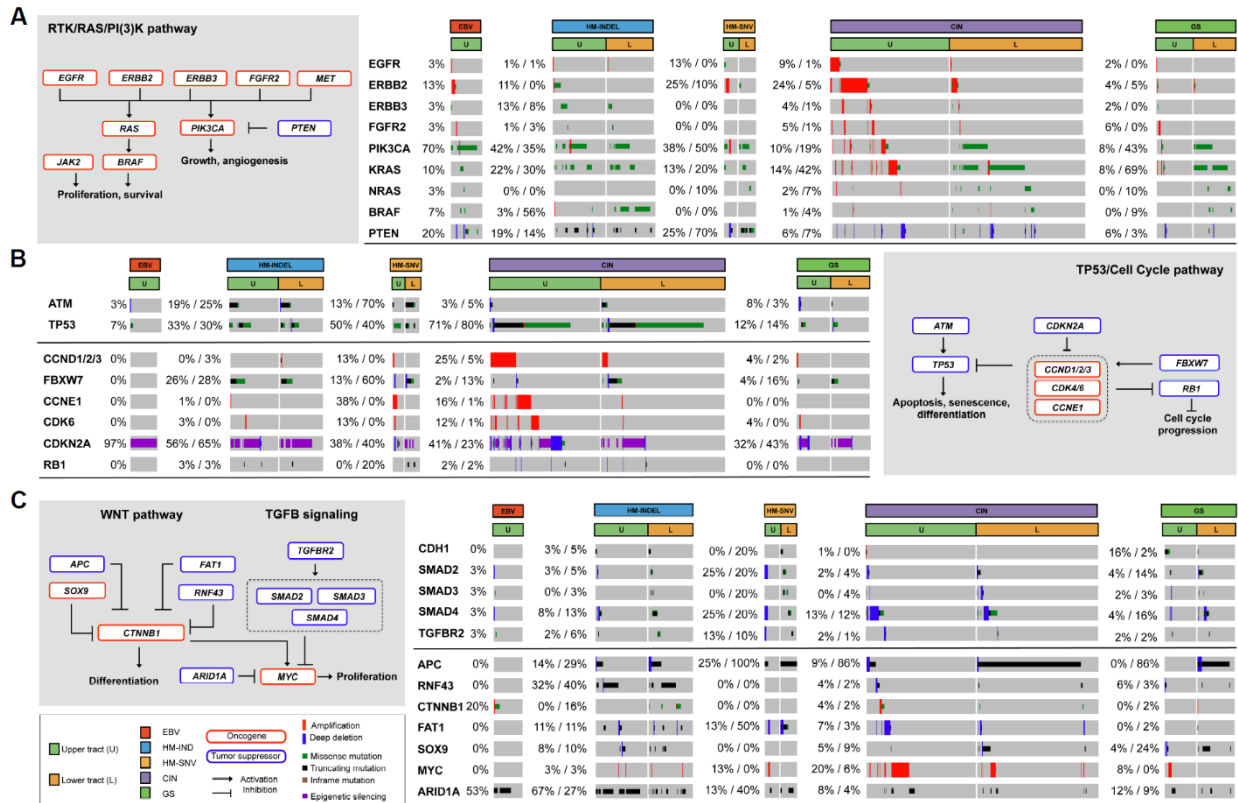


Fig. 7. Integrated molecular comparison of somatic alterations across GI molecular subtypes. A comparison of somatic disruptions across molecular subtypes as indicated by somatic mutations and copy-number alterations in select genes associated with functional pathway groups. The lines and arrows within the functional pathways show pairwise molecular interactions. Deep deletions marked in blue indicate loss of more than half of gene copies. Amplifications are marked in red. Only missense mutations reported in the COSMIC repository are included and indicated by boxes partially filled with green. Boxes partially filled with black indicate nonsense or frameshift mutations. Alteration frequencies for each gene are listed inside rounded rectangles divided by molecular subtype, with red shading denoting gene activation, and blue denoting inactivation. Percentage of somatic alteration is indicated by numbers to the left of each gene box and divided by upper (U) and lower GI (L) **(A)** Mutations and SCNAs for selected genes associated with RTK mitogen signaling pathways. **(B)** Mutations, SCNAs, and epigenetic silencing in selected genes associated with TP53/cell cycle pathways. **(C)** Mutations and SCNAs for selected genes associated with developmental pathways.

- 3) *A Pan-Cancer atlas of genomic, epigenomic and transcriptomic alterations in the TGF-beta pathway (scheduled to be submitted to Cancer Cell in early November, 2017)*
 - i) TGF-beta pathway is disrupted in gastric cancer (Fig. 8).
 - ii) The most frequently amplified TGF-beta pathway genes are *ACVR2A* (13%), *BMPR2* (10%), *SMAD4* (9%), and *SPTBN1* (8%) (Fig. 8).

- iii) The most frequently deleted TGF-beta pathway gene is *SMAD4* (6%) (Fig. 8).
- iv) Mutation frequencies are not very high, with *BMP7* having the highest frequency of 3% (Fig. 8).
- v) Hotspot mutations have been identified in *ACVR2A*, *BMPR2*, and *SMAD4* in gastric cancer (Fig. 9).
- vi) Statistically significant positive correlations have been found in gastric cancer between TGF-beta pathway activity and the hormone receptor, breast reactive, EMT, immune, hormone signaling, and PI3K/AKT pathways, whereas negative correlations have been found with cell cycle, DNA damage response and apoptosis pathways (Fig. 10).
- vii) TGF-beta pathway is highly regulated by epigenetics in gastric cancer compared to other cancers (Fig. 11). The box plot shows large dynamic range for STAD (and also DLBC) compared to others.

Stomach adenocarcinoma

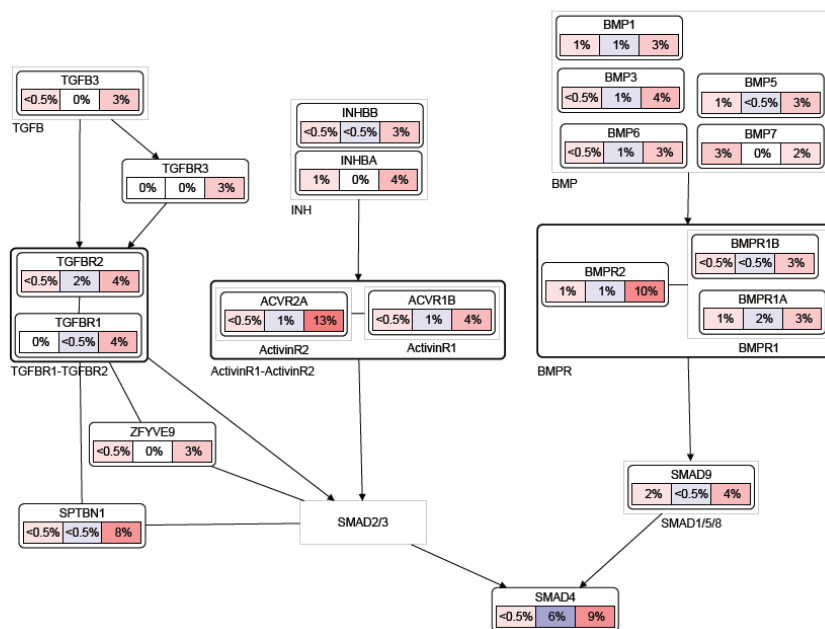
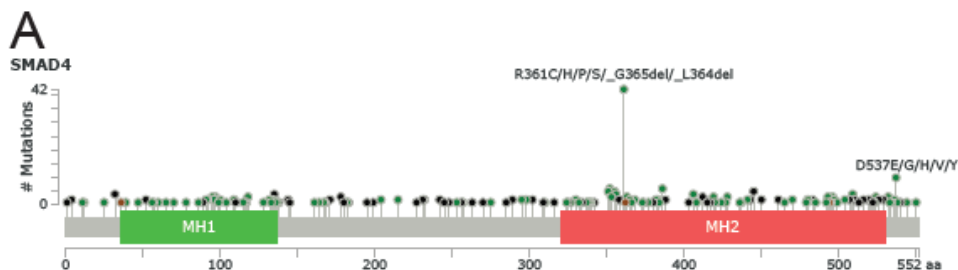


Fig. 8. Alterations in various TGF-beta pathway genes in gastric cancer. Percentages represent fraction of samples in the cohort with the given aberrations present; left box – mutations, middle box – deletions, right box – amplifications.



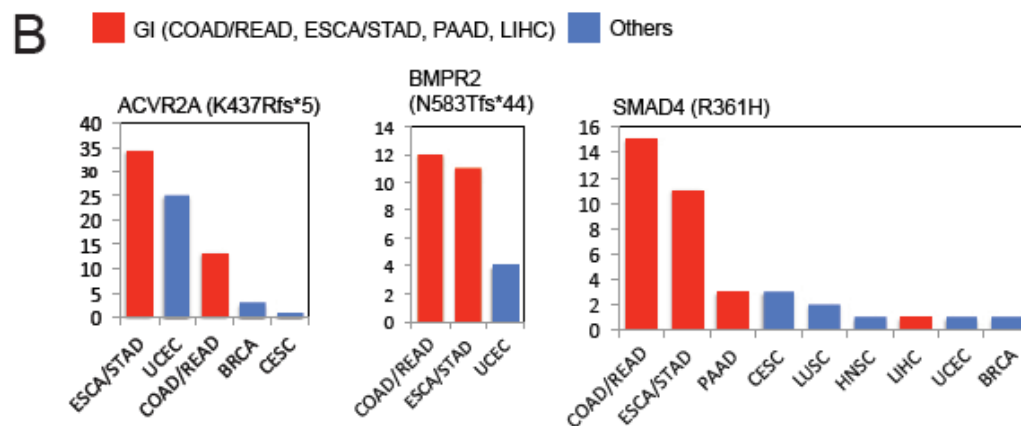
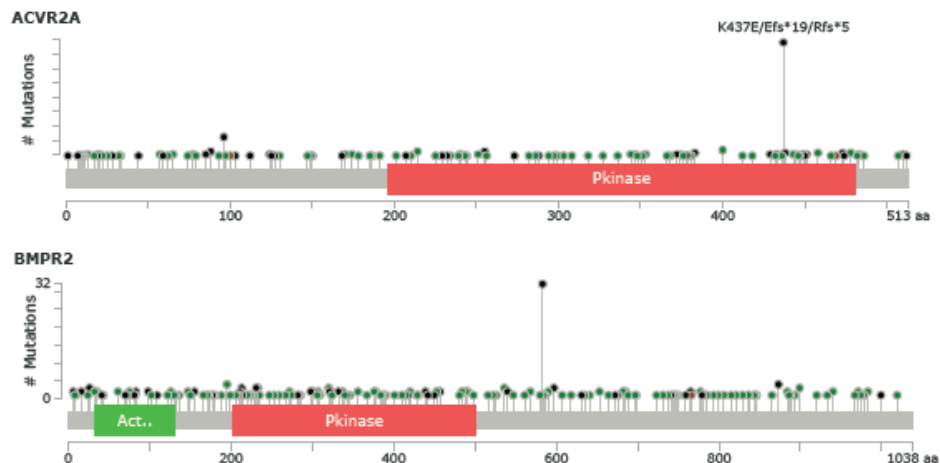


Fig. 9. (A) Lollipop plots showing mutation frequencies (y-axis) along the gene loci. Hotspot mutations have long lollipop stems. **(B)** The hotspots occur at higher frequencies in gastroesophageal and colorectal cancers than other cancers.

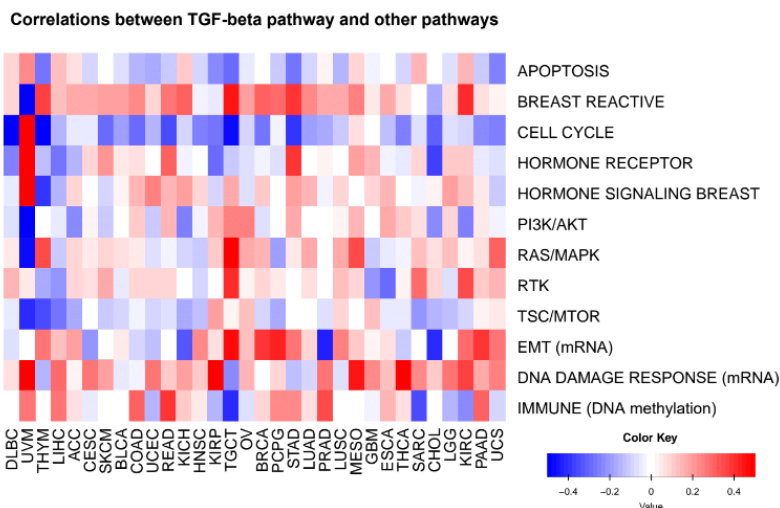


Fig. 10. Pearson's correlation coefficients between TGF-beta pathway activity and the activity of 12 other pathways (rows) across 33 disease types (columns), including gastric cancer (STAD).

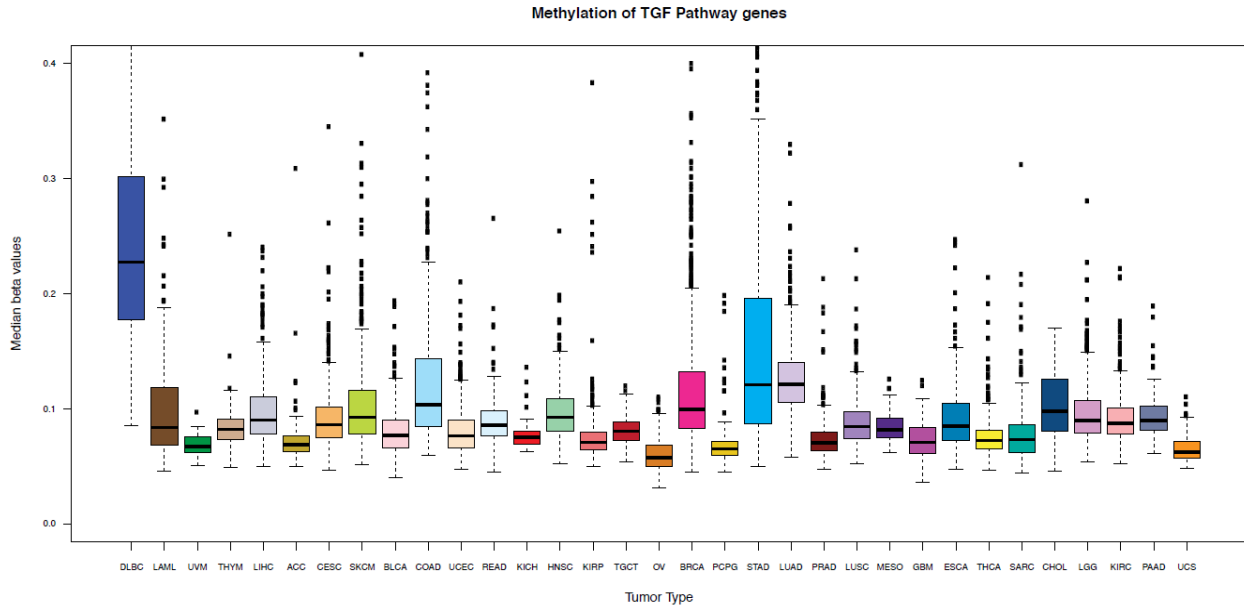


Fig. 11. Box plot showing median DNA methylation levels across the TGF-beta pathway genes (y-axis) vs. tumor types. STAD (gastric) and DLBC (B-cell lymphoma) have much larger dynamic ranges compared to other tumor types, potentially indicating a strong regulatory role of epigenetics in those diseases.

Tasks that will be performed in future under those Specific Aims

- 1) Remaining two of the three manuscripts described above will be published.
- 2) While we have studied several pathways already (TGF-beta, PI3K, RTK/RAS, mTOR, WNT, cell cycle), several other pathways will be studied in future (e.g. DNA damage response, TP53, immune etc.).
- 3) So far, we've analyzed gastric cancer in the context of other cancers, as mentioned in Specific Aims 2 and 3. Going forward, we will focus explicitly on gastric cancer, in accordance with Specific Aim 1.

What opportunities for training and professional development has the project provided?

Rehan Akbani, PhD (PI): The project has allowed Dr. Akbani several opportunities for professional development. He works with his mentor, Dr. Jaffer Ajani on a regular basis and updates him on progress. Dr. Ajani, in turn, guides Dr. Akbani's research. Dr. Ajani has also setup a larger group with about a dozen researchers who are working on gastric and other GI cancers. Dr. Akbani is part of the group and he has been afforded the opportunity to collaborate with those experts. The larger group meets on a monthly basis where researchers take turns in presenting their work and receive feedback. Dr. Akbani has also presented in that group.

Dr. Akbani is also one of the central members of The Cancer Genome Atlas (TCGA) project. TCGA funding ended in July, 2016, but the consortium continues to work on PanCanAtlas and other projects. The funding from this grant has allowed Dr. Akbani to continue collaborating with TCGA on projects related to this grant, such as Pan-GI cancers, PI3K/AKT pathway disruptions, TGF-beta pathway disruptions, etc. In some of them (e.g. TGF-beta) Dr. Akbani plays a leading role. That would not have been possible without the kind of funding provided by this grant.

Dr. Akbani presented his research in a TCGA symposium on Nov. 17, 2016, that was attended by several hundred people. As more of his research gets published, he plans to present his results

at more conferences and symposia, including a TCGA symposium planned in Washington DC in September 2018.

Shiyun Ling, PhD (postdoctoral fellow): Dr. Ling was Dr. Akbani's postdoctoral fellow. Under Dr. Akbani's guidance and supervision, Dr. Ling gained hands-on experience in performing quality control and batch effects correction of several "omics" data sets from TCGA, as stated in Major Task 4 in the approved SOW. Those data sets are now being used by TCGA for virtually all of their PanCanAtlas projects that are currently underway (approximately 2 dozen). After doing a superb job on the data, Dr. Ling accepted a permanent position as a Senior Statistical Analyst at Mount Sinai hospital in Connecticut in Spring 2017.

Apurva Hegde, MS (research assistant): Ms. Hegde was Dr. Akbani's research assistant. She performed the analysis stated in Major Task 5 under Dr. Akbani's supervision and guidance. The figures she generated have been included in the Pan-GI manuscript that has been submitted for review to Cell and Cancer Cell. She eventually accepted a permanent position as an Associate Bioinformatician at the Translational Genomics Research Institute (TGen) in Phoenix, AZ in Spring 2017.

How were the results disseminated to communities of interest?

Some of the research has already been published in renowned journals (see Products section). Other manuscripts are either currently under review, or in preparation. Besides publications, Dr. Akbani has presented the results of his research in a TCGA symposium in Nov. 2016 that was widely attended. As the research matures further, Dr. Akbani plans to attend more conferences and symposia to present his results.

What do you plan to do during the next reporting period to accomplish the goals?

- 1) Perform additional research by following the steps highlighted in the revised SOW (see Changes section).
- 2) Publish the results of the research in renowned journals.
- 3) Present the results at widely attended conferences and symposia.
- 4) Participate in cancer conferences like AACR and ASCO to improve knowledge of gastric cancer.
- 5) Collaborate with Dr. Ajani and his team of experts, and solicit regular feedback about the research.
- 6) Study the latest literature in gastric cancer to keep abreast of new developments.

IMPACT

What was the impact on the development of the principal discipline(s) of the project?

Since this is the first year of the grant and only one paper has been published so far, the impact of the research is difficult to gauge at this point. Two more publications are underway. It is expected that the papers will eventually be highly cited because, (i) they are expected to be published in the high-quality Cell family of journals, (ii) they are part of the high-profile TCGA PanCanAtlas project. Ultimately, the hope is that those and other publications that result from this grant will provide new targets for therapy in gastric cancer and increase our knowledge of the disease.

What was the impact on other disciplines?

Nothing to report.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

CHANGES/PROBLEMS

The PI is a part of the TCGA consortium. TCGA had decided to move forward with Pan-Gastrointestinal (Pan-GI) analysis last year and Dr. Akbani is a major contributor to the project. In fact, the Pan-GI paper is currently under review by Cell and Cancer Cell. In the original grant application and SOW, we had mentioned working on the Pan-GI project as part of years 2 and 3 goals. However, because of the accelerated timeline from TCGA's leadership, we ended up working on it in year 1. Year 1 goals, in turn, will be performed in year 2. Therefore, although the goals and level of effort are still the same as those listed in the original SOW, they have been reordered chronologically. The new SOW is provided in the appendix.

PRODUCTS

Publications, conference papers, and presentations

Journal publications (see appendix for details):

- 1) Zhang Y, Kwok-Shing Ng P, Kucherlapati M, Chen F, Liu Y, Tsang YH, de Velasco G, Jeong KJ, **Akbani R**, Hadjipanayis A, Pantazi A, Bristow CA, Lee E, Mahadeshwar HS, Tang J, Zhang J, Yang L, Seth S, Lee S, Ren X, Song X, Sun H, Seidman J, Luquette LJ, Xi R, Chin L, Protopopov A, Westbrook TF, Shelley CS, Choueiri TK, Ittmann M, Van Waes C, Weinstein JN, Liang H, Henske EP, Godwin AK, Park PJ, Kucherlapati R, Scott KL, Mills GB, Kwiatkowski DJ, Creighton CJ. A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer Cell*. 2017 Jun 12;31(6):820-832.e3. PMID: 28528867
- 2) Yang Liu, Nilay S. Sethi, Toshinori Hinoue Barbara G Schneider Andrew D. Cherniack, Francisco Sanchez-Vega, Jose A. Seoane, Reanne Bowlby, Mirazul Islam, Jaegil Kim, Walid Chatila, Farshad Farshidfar, **Rehan Akbani**, Rupa S. Kanchi, Charles S. Rabkin, Joseph E. Willis, Kenneth K. Wang, Shannon J. McCall, Lopa Mishra, Alexander J. Lazar, The Cancer Genome Atlas Research Network, Vésteinn Thorsson, Adam J. Bass, Peter W. Laird. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Under review by *Cell and Cancer Cell*.
- 3) Anil Korkut, Sobia Zaidi, Rupa Kanchi, Ashton C. Berger, Gordon Robertson, Lawrence N Kwong, Mike Datto, Jason Roszik, Shiyun Ling, Visweswaran Ravikumar, Ganiraju Manyam, Arvind Rao, Simon Shelley, Yuexin Liu, Zhenlin Ju, Donna Hansel, Guillermo de Velasco, Arjun Pennathur, Jesper B. Andersen, Colm J. O'Rourke, Simon Shelley, Kazu Ohshiro, Wilma Jogunoori, Nancy R. Gough, Shulin Li, Hatice Osmanbeyoglu,

Andres Houseman, Shuyun Rao, Maciej Wiznerowicz, Jian Chen, Shoujun Gu, Wencai Ma, Jiexin Zhang, Pan Tong, Andrew D. Cherniack, Chuxia Deng, Linda Resar-Smith, The Cancer Genome Atlas Research Network, Lopa Mishra, **Rehan Akbani**. A Pan-Cancer atlas of genomic, epigenomic and transcriptomic alterations in the TGF-beta pathway. Manuscript scheduled to be submitted to Cancer Cell in early November, 2017.

Conference presentations:

- 1) A Pan-Cancer analysis of TGF-beta pathway aberrations, presented by Rehan Akbani. TCGA PanCancerAtlas symposium, Houston, TX, Nov 17, 2016.

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name	Rehan Akbani
Project Role	PI
Research Identifier (e.g. ORCID ID)	
Nearest person month worked	3.6
Contribution to Project	Led the project as PI. Performed analysis. Supervised the work of others.
Funding Support	

Name	Shiyun Ling
Project Role	Postdoctoral Fellow
Research Identifier (e.g. ORCID ID)	
Nearest person month worked	6
Contribution to Project	Performed analysis under the direction of the PI
Funding Support	

Name	Apurva Hegde
Project Role	Research Assistant II
Research Identifier (e.g. ORCID ID)	
Nearest person month worked	6
Contribution to Project	Performed analysis under the direction of the PI
Funding Support	

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Yes, funding support for the PI has changed. The updated support is as follows (excluding the current grant).

Ongoing Research Support

1U24CA210950-01 (Liang and Mills) 7/1/2016-6/30/2021 1.44 calendar
NIH/NCI TCPA: an Integrated Bioinformatics Resource for Functional Cancer Proteomic Data

Major goals: To expand the scope of TCPA by adding new functionalities and datasets, and to enhance and improve its existing analytic capabilities.

1U24CA210949-01 (Weinstein, Akbani, Mills) 9/1/2016-8/31/2021 2.4 calendar
NIH/NCI

Batch effects in molecular profiling data on cancers: detection, quantitation, interpretation, and correction

Major goals: The primary goal is to analyze cancer data for batch effects for various projects specified by NCI's Center for Cancer Genomics (CCG). The data will be checked for batch effects, which will be quantified and the data corrected if needed. A secondary goal is to further enhance the batch effects analysis with new algorithms and better quality control algorithms.

5 3P30 CA016672 (Dmitrovsky) 7/1/2015-6/30/2018 0.36 calendar
NIH/NCI

Bioinformatics Shared Resources (PP-SR22)

Major goals: To assist researchers in the application of state-of-the-art methodology for the development, conduct, and analysis of studies using high-throughput technologies.

1U24 CA199461 01 (Weinstein and Broom) 9/1/2015-8/31/2020 0.48 calendar

NIH/NCI "Next Generation" Clustered Heat Maps for Fluent, Interactive Exploration of Omic Data

Major goals: Expand and enhance the capabilities of the NG-CHM system. Extend and enhance the graphical NG-CHM builder. Improve the interoperability of the NG-CHM system and integrate further with other tools, frameworks, and systems. Create additional/expanded compendia of cancer-related public datasets. Actively promote the NG-CHM system and interact with its user community.

1U24CA210950-01 (Akbani, Weinstein, Mills) 09/01/2016 – 08/31/2021 3.60 calendar
NIH/NCI

Integrated analysis of protein expression data from the Reverse Phase Protein Array (RPPA) platform

The primary goal is to analyze cancer proteomics data from the RPPA platform for various projects specified by the NCI.

Completed Research Support

NIH/NCI 5 U24 CA143883 04 Weinstein, Akbani, Mills (PI) 9/29/2009-7/31/2016

An Integrative Pipeline for Analysis & Translational Application of TCGA Data (GDAC)

The overall goal of the TGCA GDAC is to generate computational pipelines for automated integration and analysis of the data generated by the TGCA Genome Characterization Centers.

What other organizations were involved as partners?

Organization Name: The Cancer Genome Atlas (TCGA)

Location of Organization: NCI/NIH, Washington DC

Partner's contribution to the project: Collaboration (please note that funding from TCGA completed on 7/31/2017, before the reporting period started for this grant, so no funding was provided for this work by TCGA during the reporting period).

APPENDICES

Updated SOW (Chronologically reordered so Specific Aims 2 and 3 occur before Specific Aim 1. Tasks and level of effort per task remain the same. Please see “Changes” section for details.)

Specific Aim 2	Timeline (months)	Site
Major Task 4: Acquisition and Quality Control of Pan-GI data, in preparation for computational analysis		
Acquire TCGA Pan-GI data and convert them into a “standardized” format suitable for computational analysis	1	Dr. Akbani
Assess and remove (if needed) batch effects from the data, and improve the quality of the data	2	Dr. Akbani
Milestone(s) Achieved: “Cleaned up” TCGA Pan-GI data ready for computational analysis	2	
Major Task 5: Computational analysis of the Pan-GI data sets		
Cluster the data sets and study the results	3	Dr. Akbani, Dr. Ajani
Generate pathway activity scores for various pathways across multiple data types, and determine which ones are likely disrupted	4-5	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Correlate disrupted pathways across multiple data types (e.g. transcriptomic, proteomic, genomic, epigenomic) and across clinical variables (e.g. histology, stage, grade, outcome) via statistical analysis	6-7	Dr. Akbani, Dr. Ajani
Compare gastric with other Pan-GI cancers and look for similarities and differences	8	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Milestone(s) Achieved: First round computational analysis for Pan-GI cancers completed	8	
Major Task 6: Publish Pan-GI cancer results		
Discuss results with collaborators and perform any follow up analysis	9-10	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Write one or more manuscripts with input from designated mentor and collaborators	11-12	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Submit manuscript(s) and wait for reviews	13-14	Dr. Akbani

Respond to reviewers and resubmit. May repeat submission/resubmission process with multiple journals depending on where the paper(s) end up being published. Present results at conferences.	15-16	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Milestone(s) Achieved: Pan-GI manuscript(s) published	17	
Total time for Specific Aim 2	17	
Specific Aim 3 (interspersed timeline)		
Major Task 7: Identification and publication of potential therapeutic targets in gastric cancer		
Identify potential genes and/or pathways in gastric cancer for targeted therapy, using gastric data only from Aim 1	26-27	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Identify potential genes and/or pathways in gastric cancer for targeted therapy, using cross-tumor information from Pan-GI cancers from Aim 2	9-10	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Integrate results into manuscripts for Specific Aims 1 and 2. Present results at conferences.	11-17	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Milestone(s) Achieved: Potential therapeutic targets identified and published	27	
Total time for Specific Aim 3 (interspersed with other aims; not consecutive months)	27	
Specific Aim 1		
Major Task 1: Acquisition and Quality Control of gastric cancer data, in preparation for computational analysis	Months	
Acquire TCGA gastric cancer data, and in-house MD Anderson data (after procuring necessary approvals)	17.5	Dr. Akbani, Dr. Ajani
Convert all acquired data into a “standardized” format suitable for computational analysis	18	Dr. Akbani
Assess and remove (if needed) batch effects from within TCGA and within MD Anderson data, and improve the overall quality of each data set individually.	19	Dr. Akbani
Merge TCGA data with MD Anderson data, removing batch effects across both data sets.	19.5	Dr. Akbani

Re-assess the quality of the overall data and iterate back to previous steps, if needed, until data are satisfactory	20	Dr. Akbani, Dr. Ajani
Milestone(s) Achieved: “Cleaned up” gastric cancer data from TCGA and MD Anderson ready for computational analysis	20	
Major Task 2: Computational analysis of the gastric cancer data sets		
Cluster the data sets and study the results	21	Dr. Akbani, Dr. Ajani
Generate pathway activity scores for various pathways across multiple data types, and determine which ones are likely disrupted	22-23	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Correlate disrupted pathways across multiple data types (e.g. transcriptomic, proteomic, genomic, epigenomic) and across clinical variables (e.g. histology, stage, grade, outcome) via statistical analysis	24-25	Dr. Akbani, Dr. Ajani
Milestone(s) Achieved: Computational analysis for gastric cancer completed	25	
Major Task 3: Publish gastric cancer results		
Discuss results with collaborators and perform any follow up analysis	26-27	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Write one or more manuscript(s) with input from designated mentor and collaborators	28-29	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Submit manuscript(s) and wait for reviews.	30-31	Dr. Akbani
Respond to reviewers and resubmit. May repeat submission/resubmission process with multiple journals depending on where the paper(s) end up being published. Present results at conferences.	32-36	Dr. Akbani, Dr. Ajani, Dr. Weinstein, Dr. Hofstetter
Milestone(s) Achieved: Manuscript(s) published	36	
Total time for Specific Aim 1	36	

Publications

Manuscript under review by Cell and Cancer Cell. Abstract provided for reference only. Full manuscript available upon request.

Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas

Yang Liu, Nilay S. Sethi, Toshinori Hinoue Barbara G Schneider Andrew D. Cherniack, Francisco Sanchez-Vega, Jose A. Seoane, Reanne Bowlby, Mirazul Islam, Jaegil Kim, Walid Chatila, Farshad Farshidfar, **Rehan Akbani**, Rupa S. Kanchi, Charles S. Rabkin, Joseph E. Willis, Kenneth K. Wang, Shannon J. McCall, Lopa Mishra, Alexander J. Lazar, The Cancer Genome Atlas Research Network, Vésteinn Thorsson, Adam J. Bass, Peter W. Laird.

SUMMARY

We analyzed 921 adenocarcinomas of the esophagus, stomach, colon and rectum to explore the shared and distinguishing molecular characteristics of gastrointestinal tract adenocarcinomas (GIAC), probing beyond tissue-specific and anatomic boundaries. We found that hypermutated (HM) tumors were molecularly distinct regardless of cancer type, and could be subdivided into those enriched for insertions/deletions (HM-IND), representing MSI-H cases with epigenetic silencing of *MLH1* in the context of CpG Island Methylator Phenotype (CIMP), versus tumors with elevated single nucleotide variants (HM-SNV) associated with mutations in *POLE*. Tumors with chromosomal instability (CIN) displayed more diversity, with gastroesophageal adenocarcinomas harboring more fragmented genomes associated with genomic doubling and distinct mutational signatures. We identified a group of tumors in the colon and rectum lacking hypermutation and aneuploidy termed Genome Stable (GS) and enriched in DNA hypermethylation and mutations in *KRAS*, *SOX9* and *PCBP1*. This comprehensive analysis reveals molecular underpinnings of GIAC that transcend anatomic boundaries.

Manuscript to be submitted to Cancer Cell in early November, 2017. Abstract provided for reference only. Full manuscript available upon request.

A Pan-Cancer atlas of genomic, epigenomic and transcriptomic alterations in the TGF-beta pathway.

Anil Korkut, Sobia Zaidi, Rupa Kanchi, Ashton C. Berger, Gordon Robertson, Lawrence N Kwong, Mike Datto, Jason Roszik, Shiyun Ling, Visweswaran Ravikumar, Ganiraju Manyam, Arvind Rao, Simon Shelley, Yuexin Liu, Zhenlin Ju, Donna Hansel, Guillermo de Velasco, Arjun Pennathur, Jesper B. Andersen, Colm J. O'Rourke, Simon Shelley, Kazu Ohshiro, Wilma Jogunoori, Nancy R. Gough, Shulin Li, Hatice Osmanbeyoglu, Andres Houseman, Shuyun Rao, Maciej Wiznerowicz, Jian Chen, Shoujun Gu, Wencai Ma, Jiexin Zhang, Pan Tong, Andrew D. Cherniack, Chuxia Deng, Linda Resar-Smith, The Cancer Genome Atlas Research Network, Lopa Mishra, **Rehan Akbani**.

Summary

Here, we provide a multi-omic analysis of the transforming growth factor β (TGF- β)-Smad pathway in diverse human cancers. Of 9125 tumors representing 33 cancer types in The Cancer Genome Atlas, we discovered that 41 % have at least one genomic alteration in a core of 44 TGF- β pathway genes. The highest frequency occurred in gastrointestinal cancers. We identified hotspots in six genes, including those encoding ligands (*BMP5*), receptor subunits (*TGFBR2*, *AVCR2A*, *BMP2R*), and Smads (*SMAD2*, *SMAD4*). Transcriptomic analyses showed that increased or decreased TGF- β pathway activity stratified patient survival within some cancers, and tumor context was a key determinant of TGF- β function. TGF- β activity score also predicts core pathway components that are candidates for therapeutic targeting in specific cancers. Epigenetic silencing and miRNA expression, the gene repression mechanisms, provided clues to their potential role in limiting TGF- β -Smad pathway activity especially in cancers which show low pathway activity scores.

A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations

Highlights

- Multiplatform-based survey of PI3K/AKT/mTOR across over 10,000 human cancers
- Distinct classes of somatic alteration associated with greater pathway activation
- Functional interrogation of specific mutations in *PIK3CA* and *PIK3R1*
- Support for inclusion of *IDH1* and *VHL* mutations within the canonical pathway model

Authors

Yiqun Zhang, Patrick Kwok-Shing Ng, Melanie Kucherlapati, ..., Gordon B. Mills, David J. Kwiatkowski, Chad J. Creighton

Correspondence

gmills@mdanderson.org (G.B.M.),
dkwiatkowski@rics.bwh.harvard.edu (D.J.K.),
creight@bcm.edu (C.J.C.)

In Brief

Zhang et al. survey the PI3K/AKT/mTOR pathway in >10,000 human cancers across 32 types. In addition to known molecular events, some rare *PIK3CA* and *PIK3R1* mutations activate the pathway, partial copy loss of *PTEN* or *STK11* is associated with poor patient survival, and *IDH1* or *VHL* mutations can confer mTOR activity.



A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations

Yiqun Zhang,^{1,23} Patrick Kwok-Shing Ng,^{2,22,23} Melanie Kucherlapati,^{3,6} Fengju Chen,¹ Yuexin Liu,⁴ Yiu Huen Tsang,¹⁷ Guillermo de Velasco,^{5,20} Kang Jin Jeong,² Rehan Akbani,⁴ Angela Hadjipanayis,^{3,6} Angeliki Pantazi,^{3,7} Christopher A. Bristow,⁸ Eunjung Lee,^{3,6} Harshad S. Mahadeshwar,⁸ Jiabin Tang,⁸ Jianhua Zhang,⁸ Lixing Yang,⁹ Sahil Seth,⁸ Semin Lee,⁹ Xiaojia Ren,^{3,7} Xingzhi Song,⁸ Huandong Sun,⁸ Jonathan Seidman,³ Lovelace J. Luquette,⁹

(Author list continued on next page)

¹Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

²Department of Systems Biology, University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA

³Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

⁴Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

⁶Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA

⁷KEW Inc., Cambridge, MA 02139, USA

⁸Department of Genomic Medicine, Institute for Applied Cancer Science, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁹Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

¹⁰The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology, Harvard University Cambridge, Cambridge, MA 02142, USA

¹¹Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA

¹²Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX 77030, USA

(Affiliations continued on next page)

SUMMARY

Molecular alterations involving the PI3K/AKT/mTOR pathway (including mutation, copy number, protein, or RNA) were examined across 11,219 human cancers representing 32 major types. Within specific mutated genes, frequency, mutation hotspot residues, in silico predictions, and functional assays were all informative in distinguishing the subset of genetic variants more likely to have functional relevance. Multiple oncogenic pathways including PI3K/AKT/mTOR converged on similar sets of downstream transcriptional targets. In addition to mutation, structural variations and partial copy losses involving *PTEN* and *STK11* showed evidence for having functional relevance. A substantial fraction of cancers showed high mTOR pathway activity without an associated canonical genetic or genomic alteration, including cancers harboring *IDH1* or *VHL* mutations, suggesting multiple mechanisms for pathway activation.

INTRODUCTION

The phosphatidylinositol 3-kinase (PI3K)/AKT/mammalian target of rapamycin (mTOR) signaling pathway is one of the main growth regulatory pathways in both normal cells and cancer (Hennessy et al., 2005; Mayer and Arteaga, 2016). This growth

pathway begins with class IA PI3Ks, which are heterodimers consisting of p110 catalytic and p85 regulatory subunits. Growth factor receptor tyrosine kinases (RTKs) activate PI3K through phosphorylation of adaptor proteins such as IRS1/IRS2 (Engelman et al., 2006). These adaptor proteins bind the amino-terminal domain of the PI3K p85 regulatory subunits through YXXM

Significance

Our current model of the PI3K/AKT/mTOR pathway has largely been derived from experimental systems. The Cancer Genome Atlas (TCGA) pan-cancer cohort represents an opportunity to explore these pathway relationships in the setting of human cancer. Cause-and-effect relationships embodied by the pathway model can manifest as correlations in human disease. Integration of genomic with proteomic data may benefit personalized and precision medicine approaches in helping to assess variants for potential clinical relevance. Manifestation of pathways at the transcription level is distinct from manifestation at the phospho-protein level, highlighting the importance of proteomic approaches. Over time, previously un-realized or underappreciated members or connections may be incorporated into the standard pathway model, where TCGA data may aid in the process of discovery or confirmation.

Ruibin Xi,⁹ Lynda Chin,^{8,10} Alexei Protopopov,^{8,7} Thomas F. Westbrook,^{1,17,21} Carl Simon Shelley,¹¹ Toni K. Choueiri,⁵ Michael Ittmann,¹² Carter Van Waes,¹³ John N. Weinstein,⁴ Han Liang,^{4,2} Elizabeth P. Henske,^{14,15} Andrew K. Godwin,¹⁶ Peter J. Park,^{9,6} Raju Kucherlapati,^{3,6} Kenneth L. Scott,¹⁷ Gordon B. Mills,^{2,*} David J. Kwiatkowski,^{14,15,*} and Chad J. Creighton^{1,4,18,19,24,*}

¹³Tumor Biology Section, Head and Neck Surgery Branch, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892, USA

¹⁴The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology, Harvard University, Cambridge, MA 02142, USA

¹⁵Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA

¹⁶Department of Pathology & Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66160, USA

¹⁷Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

¹⁸Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA

¹⁹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

²⁰Department of Medical Oncology, University Hospital 12 de Octubre, Madrid 28041, Spain

²¹Department of Biochemistry & Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

²²Sheikh Khalifa Bin Zayed Al Nahyan Institute for Personalized Cancer Therapy, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²³These authors contributed equally

²⁴Lead Contact

*Correspondence: g mills@mdanderson.org (G.B.M.), dkwiatkowski@rics.bwh.harvard.edu (D.J.K.), creight@bcm.edu (C.J.C.)
<http://dx.doi.org/10.1016/j.ccell.2017.04.013>

motifs, to reverse its inhibition of the p110 catalytic subunit, and leads to movement of the p85–p110 heterodimer to the cell membrane where p110 can phosphorylate phosphatidylinositol-4,5-bisphosphate (PIP2) to generate phosphatidylinositol-3,4,5-trisphosphate (PIP3). RAS family members can also activate PI3K (Mayer and Arteaga, 2016). The primary negative regulator of PI3K activation is the phosphatase PTEN, which dephosphorylates PIP3 at the 3' position (Keniry and Parsons, 2008), with a secondary negative regulator being INPP4B (inositol polyphosphate-4-phosphatase type II B). PIP3 recruits several pleckstrin homology domain-containing proteins to the membrane, including AKT and PDK1. AKT is phosphorylated at Thr308 by PDK1 and at Ser473 by mTOR complex 2 (mTORC2), which increases its kinase activity. AKT directly and indirectly phosphorylates many downstream proteins, including the GSKs, p27KIP1, FoxO transcription factors, MDM2, and BAD, to enhance cell survival and growth (Manning and Cantley, 2007). Furthermore, AKT phosphorylates TSC2 at multiple sites, to inhibit the GTPase-activating protein function of the TSC protein complex (consisting of TSC1, TSC2, and TBC1D7) toward Rheb, a RAS family member (Dibble and Manning, 2013; Laplante and Sabatini, 2012). Rheb-GTP binds to mTOR complex 1 (mTORC1) to activate its kinase activity toward the S6Ks, 4E-BP1, and other substrates, leading to enhancement of multiple anabolic biosynthetic pathways that enable production of the building blocks (e.g., nucleotides) and macromolecules (e.g., ribosomes) required for cell size increase and mitosis (Dibble and Manning, 2013).

Multiple genetic events have been described that lead to activation of the PI3K/AKT/mTOR pathway in cancer (Thorpe et al., 2015). Activating mutations in *PIK3CA*, which encodes the PI3K p110a catalytic subunit, are common in many cancer types (Samuels et al., 2004; Thorpe et al., 2015). There are highly focal hotspots of mutation in *PIK3CA*, E542, and E545 in the helical domain, and H1047 and G1049 in the kinase domain, which activate the kinase through different mechanisms. Other PI3K p110 isoforms are rarely mutated in cancer overall, but *PIK3CA* and *PIK3CB*, as well as the class II PI3K

PIK3C2B, are all amplified in one or more cancer types (Thorpe et al., 2015). *PIK3R1*, and less commonly *PIK3R2*, which encode the p85 α and p85 β regulatory subunits of PI3K, are commonly mutated, resulting in reduced ability to inhibit PI3K p110a (Cheung et al., 2011; Thorpe et al., 2015). *PTEN* is subject to both genomic deletion and small point mutations that inactivate its function, and is one of the most commonly mutated cancer genes overall (Keniry and Parsons, 2008). *AKT1* is occasionally activated by mutation at a single site, E17K (Carpten et al., 2007). Inactivating mutations in both *TSC1* and *TSC2* have been identified in cancer at low frequency (Hornigold et al., 1999), as well as activating mutations in *MTOR* (Grabiner et al., 2014). *RHEB* mutations are rare but focal at Y35, suggesting a driver effect.

With the recent conclusion of the data generation phase of The Cancer Genome Atlas (TCGA), there is opportunity for systematic analyses of the entire TCGA pan-cancer cohort, including analyses focusing on specific oncogenic pathways. The aim of our study was to comprehensively examine the entire PI3K/AKT/mTOR pathway and its components in over 10,000 human cancers and 32 cancer types profiled by TCGA, using multiple molecular profiling platforms, including proteomics.

RESULTS

Proteomic Analysis of the PI3K/AKT/mTOR Pathway

Our study involved 11,219 human cancer cases representing 32 different major types, for which TCGA generated data on one or more of the following molecular characterization platforms (Table S1): whole-exome sequencing (WES, n = 10,224 cases), whole-genome sequencing data (WGS, n = 1,363), somatic DNA copy by SNP array (n = 10,845), RNA sequencing (n = 10,224), and reverse-phase protein array (RPPA). We used the RPPA proteomic platform to analyze 7,663 patient samples from 31 cancer types (with no data available for AML patients). The RPPA dataset comprised 225 high-quality antibodies that target 166 total proteins and 56 phosphorylated proteins. In

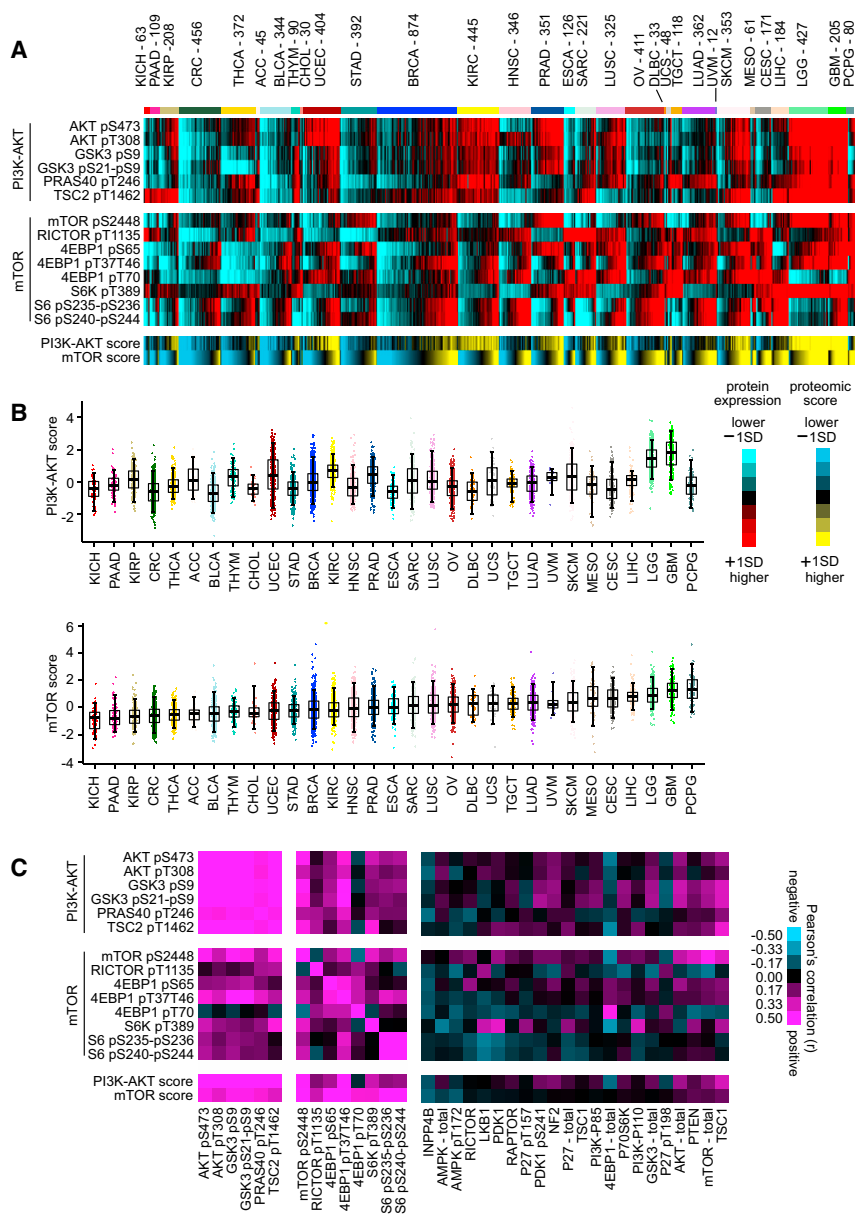


Figure 1. Proteomic Signatures of PI3K/AKT and mTOR across Human Cancers

(A) Heatmap of RPPA features considered core to either PI3K/AKT or mTOR pathways across 7,663 cancers. Red, higher expression (values normalized to SDs from the median across all cancers); blue, lower expression. PI3K/AKT and mTOR features were each summarized into pathway activity scores for each tumor profile (yellow, higher inferred activity; blue, lower activity; bright yellow/blue denotes change of 1 SD or SD, from the median). Cancer types (denoted by TCGA project name) are ordered by low to high average mTOR pathway score.

(B) Boxplots of PI3K/AKT (top) and mTOR (bottom) pathway activities scores, as inferred using RPPA data. Boxplots represent 5%, 25%, 50%, 75%, and 95%.

(C) Pearson's correlations between RPPA features across all cancers, involving features core to PI3K/AKT or mTOR pathways, as well as involving features representing proteins that may act peripherally upon either pathway. See also [Figure S1](#) and [Tables S1](#) and [S2](#).

On average, mTOR scores differed by tumor lineage, with, for example KICH (kidney chromophobe) tumors showing the lowest levels of mTOR activity, and with PCPG (pheochromocytoma and paraganglioma) showing the highest levels (followed by glioblastoma multiforme and brain lower grade glioma (LGG), or glioblastoma and LGG, respectively); at the same time, within each tumor type a wide range of activity levels were evident ([Figure 1B](#)). Across tumor profiles, PI3K/AKT and mTOR activity scores were highly significantly correlated (Pearson's $r = 0.50$, $p \sim 0$), although many cancer cases showed high mTOR activity but low PI3K/AKT activity or vice versa ([Figures 1A](#) and [1B](#)), indicative of a certain degree of decoupling between the two pathway branches. Individual members of the

this study, we carried out data normalization and batch correction to allow for direct comparisons between different cancer types. In general, mRNA levels were significantly correlated with protein levels, but strong correlations between mRNA and phospho-protein levels involving PI3K/AKT/mTOR pathway members were not observed ([Figure S1A](#)). In this study, we regarded mTOR signaling as a separate pathway from PI3K/AKT, where the former integrates information from the PI3K/AKT, Ras/MAPK, and LKB1/AMPK pathways ([Laplante and Sabatini, 2012](#)). Following previous studies ([Akbani et al., 2014](#)), we developed pathway signatures for both PI3K/AKT and mTOR components, based on member proteins selected by literature review, as a means of assessing the overall level of pathway activity given the variations of individual members.

For each tumor, RPPA signatures for PI3K/AKT and mTOR were summarized into activity scores ([Figure 1A](#) and [Table S2](#)).

PI3K/AKT signature were strongly correlated in protein expression with each other across cancers ([Figure 1C](#)). mTOR pathway-related members were also highly inter-correlated ([Figure 1C](#)), with distinct clusters involving 4EBP1- and S6-related features, respectively, and with phospho-RICTOR negatively correlated with phospho-mTOR ($r = -0.14$, $p < 1E-30$). Other protein features strongly correlated with PI3K/AKT/mTOR signaling included members of the MAP Kinase pathway ([Figures S1B](#) and [S1C](#)). When considering a number of additional RPPA features for proteins understood to act peripherally on PI3K/AKT or mTOR signaling, these tended to show weaker correlations with PI3K/AKT and mTOR features ([Figure 1C](#)). INPP4B and AMPK were negatively correlated with mTOR activity as expected, while within subsets of tumors other proteins would presumably have pathway-related roles that may not be reflected in more global analyses.

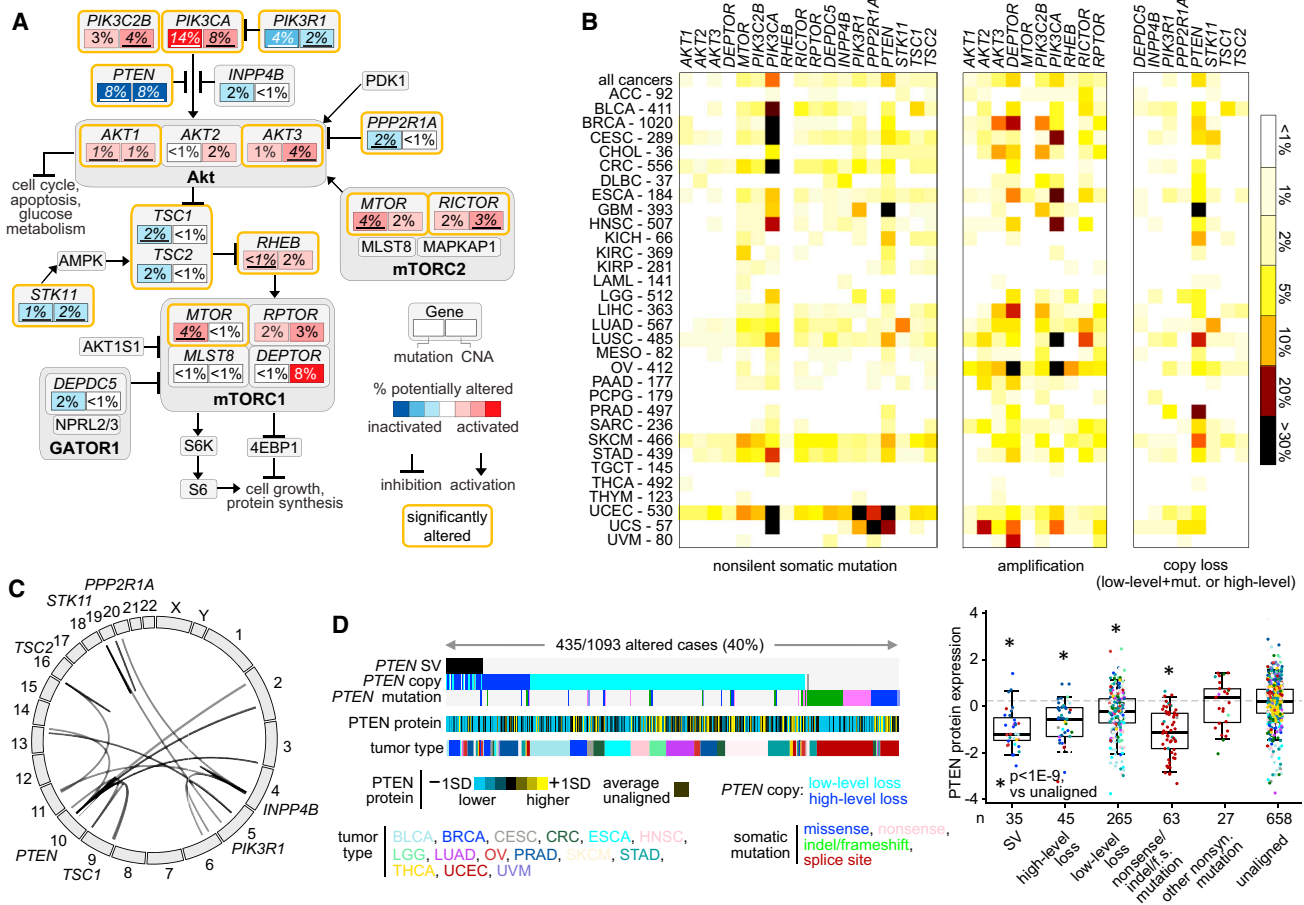


Figure 2. Somatic Mutations and DNA Copy and Structural Alterations Involving Components of the PI3K/AKT/mTOR Pathway across Human Cancers

(A) Diagram of somatic mutation and copy-number alteration (CNA) frequencies involving components of the PI3K/AKT/mTOR pathway. Key genes (with significant or sizable frequencies of alteration) are indicated by rectangles, with the percentages of somatic mutations and CNAs shown in the left and right portions of each rectangle, respectively. Significantly altered genes (from Chang et al., 2016; Kandoth et al., 2013; Lawrence et al., 2014; Zack et al., 2013; percentages representing significant alterations are underlined) are bounded by orange lines. Red, potentially activating genetic alterations; blue, potentially inactivating genetic alterations. Copy loss represents either “high-level” deletion (approximating homozygous deletion) or mutation in combination with “low-level” deletion (partial loss).

(B) By cancer type, percentages of somatic mutation or copy alteration for each indicated gene. Amplification denotes “high-level” copy gain. Numbers of cases denote representation on WES data platform.

(C) Genomic rearrangements (represented in circos plot) involving *PTEN*, *INPP4B*, *STK11*, *TSC1*, *TSC2*, *PIK3R1*, or *PPP2R1A*, based on analysis of 1,363 cases with WGS data.

(D) Left: alterations involving *PTEN* (somatic mutation, copy alteration, structural variation, or SV) found in the set of 1,093 cancers cases having both WGS and RPPA data available (protein values normalized to SDs, or SDs, from the median). Right: boxplot of *PTEN* protein expression by alteration class. Boxplots represent 5%, 25%, 50%, 75%, and 95%. p Values by t test on log-transformed values. See also Figure S2 and Tables S3 and S4.

Somatic DNA Alterations Involving the PI3K/AKT/mTOR Pathway

We examined gene mutations (using WES, n = 10,224 cases) and somatic DNA copy alterations (by SNP 6.0 arrays, n = 10,845 cases), focusing on genes in the canonical PI3K/AKT/mTOR pathway (Figure 2A and Table S3). Frequencies of somatic alteration for key genes in the pathway were tabulated across all cancers as well as within each cancer type according to TCGA project (Figure 2B and Table S4). A number of genes in the pathway were found significantly mutated or copy altered in pan-cancer analyses (Chang et al., 2016; Kandoth et al., 2013; Lawrence et al., 2014; Zack et al., 2013), including *PIK3CA*

(14% mutated across all cancers; 6% amplified), *PTEN* (9% mutated; 7% deletion or two-hit loss), *PIK3R1* (4% mutated), *PPP2R1A* (2% mutated), *AKT1* (1% mutated), *AKT1* (3% amplified), *TSC1* (2% mutated), *STK11* (2% mutated; 1% deletion or two-hit loss), *RICTOR* (3% amplified), and *MTOR* (4% amplified). With the notable exception of *AKT3*, copy number alterations of PI3K/AKT/mTOR pathway member genes were highly correlated with their mRNA expression (Figure S2A). When overlaid with mutation frequency data from human tumors, the model of the PI3K/AKT/mTOR pathway (Figure 2A) can indicate which pathway members or interactions may be most relevant in the context of cancer. However, even genes with a low frequency

of DNA alterations (e.g., *AKT1*, *MAPKAP1*, *MLST8*, *PDK1*) may be critical in individual cancer cases or in specific cancer types or subtypes not included here, in which they may be more commonly altered.

Genomic rearrangements represent another class of somatic alterations impacting gene function. Out of 1,363 cases with WGS data available (1,218 by low-pass sequencing), 63 cases (~5%) harbored a rearrangement within pathway suppressor genes *PTEN* (39 cases), *INPP4B* (14), *STK11* (5), *TSC1* (2), *TSC2* (2), *PIK3R1* (2), or *PPP2R1A* (2) (Figure 2C). By structural variation (SV), copy loss (partial or total), or mutation, *PTEN* was found altered in 40% of cancers with both RPPA and WGS data, with *PTEN* protein expression most impacted in tumors with SV, homozygous loss, or nonsense/indel/frameshift mutations (Figure 2D). In addition to *PTEN*, SVs within *STK11* and *TSC1* were also associated with decreased expression (Figure S2B). Furthermore, high- and low-level copy number loss for several pathway genes were strongly correlated with reduced mRNA levels (Figure S2B), and 20 cases harbored candidate gene fusions involving *PIK3CA*, *AKT1*, *AKT2*, *AKT3*, or *MTOR* (Figure S2C and Table S3).

Recurrently Mutated Residues in Key Genes Associated with Protein Activation

A large proportion of mutations identified in driver genes that activate PI3K/AKT/mTOR are of low occurrence, highlighting the need to functionally annotate the long tail of infrequent mutations present in heterogeneous cancers (Dogruluk et al., 2015). For example, *PIK3CA* is the gene most commonly activated by mutation in the cancer genome, with mutations being most frequent at positions E542, E545, and H1047 (Figure 3A); on the other hand, 13% of *PIK3CA* mutations observed occurred in a single case and showed no significant pattern of occurrence. Somatic copy alteration represents another potential mechanism for altering gene function where, for example, amplification of *PIK3CA* impacts p110 α protein expression (Figure 3B). Previous pan-cancer sequence analyses (Chang et al., 2016) have identified recurrent mutational hotspots, where such hotspots would presumably have greater impact on protein function. In the case of *PIK3CA*, 73% of somatic, nonsilent mutation variants identified in TCGA pan-cancer cohort involved a hotspot residue as identified by Chang et al., while 13% of *PIK3R1* mutations and 7% of *MTOR* mutations involved a hotspot residue (Figure 3C). In addition, algorithms such as Mutation Assessor (Reva et al., 2011) have predicted the likely functional impact of somatic mutation, e.g., based on evolutionary conservation of the affected amino acid in protein homologs.

As the above genes, as well as *PTEN*, presumably act upon AKT (Figure 2A), phospho-protein expression of AKT was examined in relation to tumor groups as defined by somatic alteration of a key gene (Figure 3D). For each gene considered, mutations were separated on the basis of whether or not a prediction of mutation functionality could be made (by residue hotspot, by Mutation Assessor, by manual literature review, or by nonsense/frameshift/indel involving *PTEN* or *PIK3R1*). For each of the genes considered (*AKT1*, *MTOR*, *PIK3CA*, *PIK3R1*, and *PTEN*), tumors harboring mutations that were predicted to have functional effects had elevated phospho-AKT levels on average, compared with tumors that did not harbor an alteration; in addi-

tion, tumors with mutations not predicted to be functional showed either a lesser effect or no significant effect on phospho-AKT. *PTEN* copy losses were also associated with AKT activation, while, interestingly, *PIK3CA* amplifications and copy alterations involving other specific genes (Figure S3) were not.

In addition to analysis of significantly mutated residues and of phospho-protein expression, functional studies using cell lines represents another way to annotate mutations in terms of their oncogenic potential. Using MCF10A and Ba/F3 cells, 69 different nonsilent *PIK3CA* mutation variants were functionally assessed in vitro for their activating potential (Figures 4A and S4; Table S5). Most variants tested showed some level of functionality (from weak to strong) in at least one of the two cell lines, while 14 variants showed no functional effects and two showed inhibitory or inactivating effects. The degree of growth activation varied considerably, with the three highly recurrent *PIK3CA* site (E542, E545, and H1047) mutants showing some of the highest degrees of activity in this assay. In another experiment, 35 different nonsilent *PIK3R1* mutation variants were functionally interrogated in Ba/F3 cells (Figure 4B). When the results of the functional studies were aligned with data from TCGA, a significant trend was observed for both *PIK3CA* and *PIK3R1*, whereby variants that were associated with functionality in vitro had a higher frequency of occurrence in human tumors (Figure 4C), suggesting that natural selection favored tumor development for those variants with greater functional effects. Most variants showing some functionality also had higher phospho-AKT on average, compared with tumors with the corresponding wild-type gene (Figures 4A and 4B), although variants associated with higher phospho-AKT were not necessarily associated with higher phospho-TSC2 (downstream in the pathway from AKT).

Transcriptomic Analysis of PI3K/AKT/mTOR Pathway

Signaling pathways that influence cell growth transduce signals to the nucleus, leading to activation or deactivation of the transcription of specific genes (Hanahan and Weinberg, 2000). Previously, we had defined a PI3K/AKT/mTOR transcriptional (mRNA) signature, based on the set of genes either induced or repressed by PI3K or mTOR inhibitors (Creighton et al., 2010). We applied this signature to the Library of Integrated Network-based Cellular Signatures (LINCS) database (Duan et al., 2014) of perturbational expression profiles across multiple cell and perturbation types. In the LINCS L1000 expression dataset (consisting of ~1,000 genes), the PI3K/AKT/mTOR mRNA signature was inversely associated with the transcriptional responses of cell lines to PI3K/AKT/mTOR inhibitors (Figure 5A). We evaluated the signature against the LINCS expression profiles of cells treated with short hairpin RNA (shRNAs) for ~6,000 different genes; knock down of pathway effectors (e.g., *MTOR* or *RPTOR*) resulted in gene signature patterns inversely correlated to those of our PI3K/AKT/mTOR signature, while knock down of pathway suppressors (e.g., *PTEN* or *INPP4B*) resulted in signature patterns positively correlated with those of our signature (Figure 5B and Table S6). Notably, knock down of *MYC* and *KRAS* also suppressed the PI3K/AKT/mTOR signature; furthermore, when scoring TCGA pan-cancer mRNA profiles for pre-defined signatures of PI3K/AKT/mTOR, *MYC*, and k-ras, cancers scoring high for PI3K/AKT/mTOR also tended to score high for *MYC* and k-ras

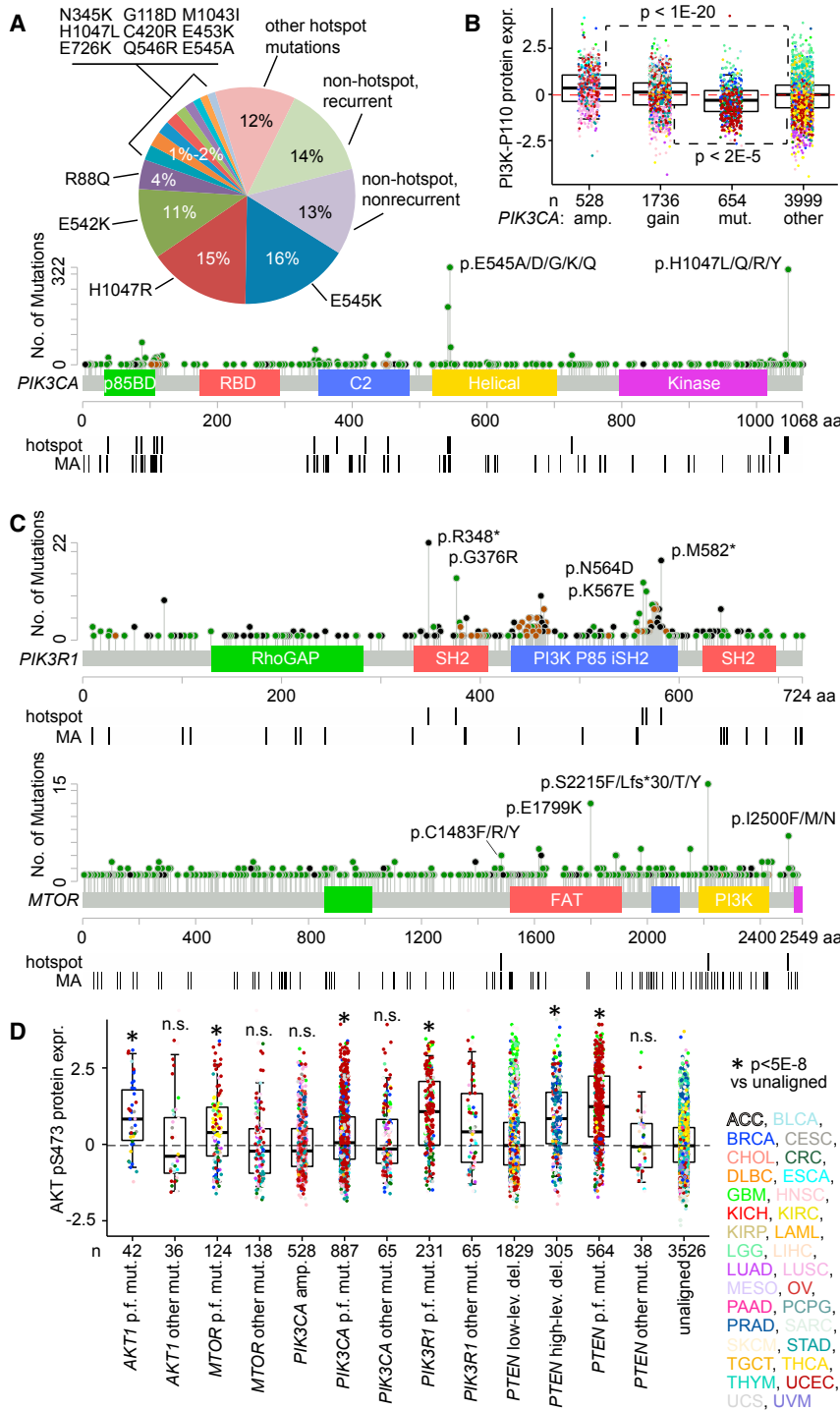


Figure 3. Distributions of Mutations in Key PI3K/AKT/mTOR Pathway Genes and Association with Protein Activation

(A) *PIK3CA* nonsilent, somatic variant frequencies and distribution across domain-annotated p110 α protein structure. “Recurrent” denotes mutation event observed in two or more tumor cases. “Hot-spot” denotes recurrently mutated residues as identified by pan-cancer sequence analyses (Chang et al., 2016). “MA” denotes “medium” or “strong” functional prediction by Mutation Assessor algorithm (Reva et al., 2011).

(B) Boxplot of p110 α expression by *PIK3CA* alteration class (gene amplification, gain of one to two copies, mutation, or none of the above, i.e. “unaligned”). p Values by t test on log-transformed values.

(C) Distributions of nonsilent and somatic variants within *PIK3R1* (top) and *MTOR* (bottom) across their respective domain-annotated protein structures.

(D) Boxplot of AKT pS473 phospho-protein expression by mutation (mut.) or copy alteration class, with the unaligned cases having none of the listed alteration types. p.f., predicted functional mutations (by hotspot, Mutation Assessor analysis, literature review, or nonsense/frameshift/indel involving *PTEN* or *PIK3R1*); amp., high-level gene amplification; low-lev. and high-lev., low- and high-level copy deletions, respectively. p Values by t test on log-transformed values. n.s., not significant ($p > 0.05$). Boxplots represent 5%, 25%, 50%, 75%, and 95%. Points in boxplots are colored according to tumor type as defined by TCGA project as indicated in (D). See also Figure S3.

As a means of identifying a transcriptional signature associated with the PI3K/AKT/mTOR pathway, we examined datasets from Garnett et al. (2012), which included both gene expression and drug sensitivity data for 131 drugs on a set of 594 human cancer cell lines. To derive gene expression correlates of sensitivity to pathway inhibition, half maximal inhibitory concentration values for 11 different inhibitors to PI3K/AKT/MTOR were normalized and averaged to obtain a single drug sensitivity score across cell lines. After correcting for expression differences specific to tumor type, 146 genes were significantly associated ($p < 0.01$, generalized linear model) with pathway inhibitor sensitivity

(Figures 5C, S5A, and S5B), suggesting that multiple oncogenic signaling pathways may converge on similar sets of transcriptional targets. The above mRNA signatures would represent more than cell proliferation processes, given how the signatures were originally derived (Creighton et al., 2010), the lack of cell-cycle regulators in the top LINCS shRNA results (Figure 5B), and the signature association with key alterations in human tumors (Figure 5C).

(Figure 5D and Table S6). Across cell lines, this inhibitor sensitivity signature correlated significantly with PI3K/AKT phospho-protein levels (Figure 5E), but showed little overlap with the above CMAP signature (from Figure 5A). Furthermore, when scoring TCGA pan-cancer mRNA profiles for the above signatures, tumors that scored high for the inhibitor signature tended to score low for the CMAP signature and vice versa, and PI3K/AKT proteomic score (but not mTOR score) was

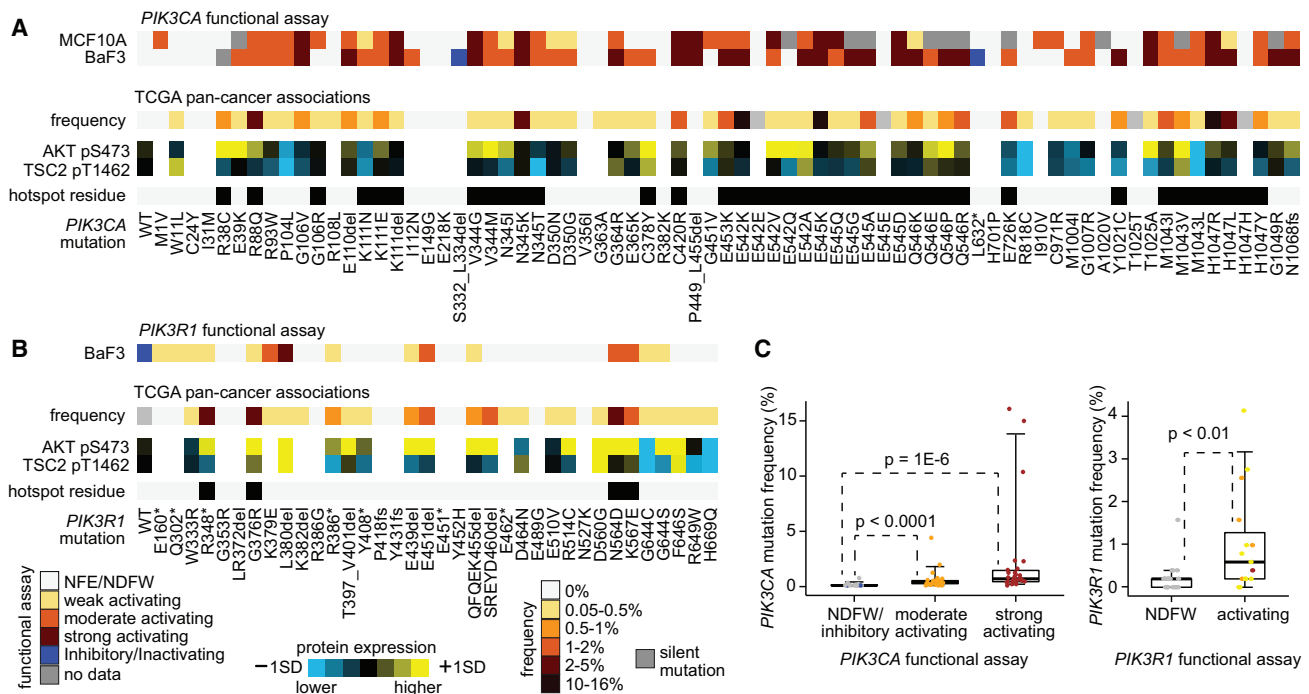


Figure 4. Functional Assessment of Specific *PIK3CA* and *PIK3R1* Variants by Cell Line Viability Assays

(A and B) Ba/F3 or MCF-10A cells were transfected with wild-type (WT) or indicated mutant cDNA of *PIK3CA* (A) or *PIK3R1* (B) then cultured for 4 weeks and harvested for viability assay. The extent of functionality conferred by the variant is indicated by colorgram. NFE/NDFW, no functional effect/no difference from wild-type. For the mutant variants assessed, corresponding human cancer data from TCGA are shown, including frequency of the variant (relative to other variants found for the same gene) and average protein expression for AKT pS473 and TSC2 pT1462. Hotspot residue, from Chang et al. (2016).

(C) For *PIK3CA* (left) and *PIK3R1* (right), boxplots of variant frequency in TCGA human tumors (relative to other variants found for the same gene) by functional assays results. p Values by Mann-Whitney U test. Boxplots represent 5%, 25%, 50%, 75%, and 95%. See also Figure S4 and Table S5.

again highly correlated with the inhibitor signature score (Figures 5E, S5C, and S5D).

Molecular Correlates of Patient Survival Involving PI3K/AKT/mTOR Pathway Components

Molecular correlates of cancer patient survival can offer insights into the pathways and processes underlying more aggressive disease (The Cancer Genome Atlas Research Network, 2013). For specific cancer types (e.g., breast and lung adenocarcinoma), the PI3K/AKT/mTOR pathway has been associated with aggressive disease (The Cancer Genome Atlas Network, 2012; The Cancer Genome Atlas Research Network, 2013). In this present study, we sought to define survival correlates in pan-cancer analyses, leveraging the large numbers of patients available (these numbers helping to balance the relatively short patient follow-up times that characterize a number of individual TCGA projects). As some cancer types are inherently more aggressive than others (Hoadley et al., 2014), we carried two separate tests for each molecular feature examined: an “uncorrected” test across all cancers regardless of type and a “corrected” test incorporating cancer type (by TCGA project) as a covariate. Features more strongly associated with an aggressive cancer type but having a survival association that was not independent of cancer type (*PIK3CA* mutation, for example, Hoadley et al., 2014) may show significance for the uncorrected but not the corrected survival test.

Numerous protein expression or genomic alteration features involving PI3K/AKT/mTOR pathway members were significantly associated with patient outcome in pan-cancer analyses (Figure 6A), a number of these features remaining significant after correcting for cancer type. Features significantly associated with worse patient outcome, independent of cancer type, included *STK11* mutation, *STK11* copy loss, *PTEN* copy loss, *PIK3CA* amplification, and higher phospho-4EBP1 expression. Focusing on *PTEN* and *STK11* copy alterations, these features were found significant within several individual cancer types, with the aggregated patterns across cancer types denoting pan-cancer significance (Figure 6B). Interestingly, for both *PTEN* and *STK11*, low-level deletion (approximating partial copy loss) but not high-level deletion (approximating total loss) was associated with significantly worse outcome compared with wild-type (Figures 6C and 6D); loss of one copy combined with somatic mutation of the other copy was associated with the poorest outcome. For both *PTEN* and *STK11*, neither high-level deletion nor mutation without copy loss could be associated with worse outcome, where in this instance, survival differences by tumor type were a likely confounder (e.g., 65% of the *PTEN* mutation with no copy alteration group were UCEC, or uterine corpus endometrial carcinoma, cases). As a group, gene transcription targets of the PI3K/AKT/mTOR pathway (based on the signature described in Figure 5A) were also associated with worse patient outcome (Figure 6E).

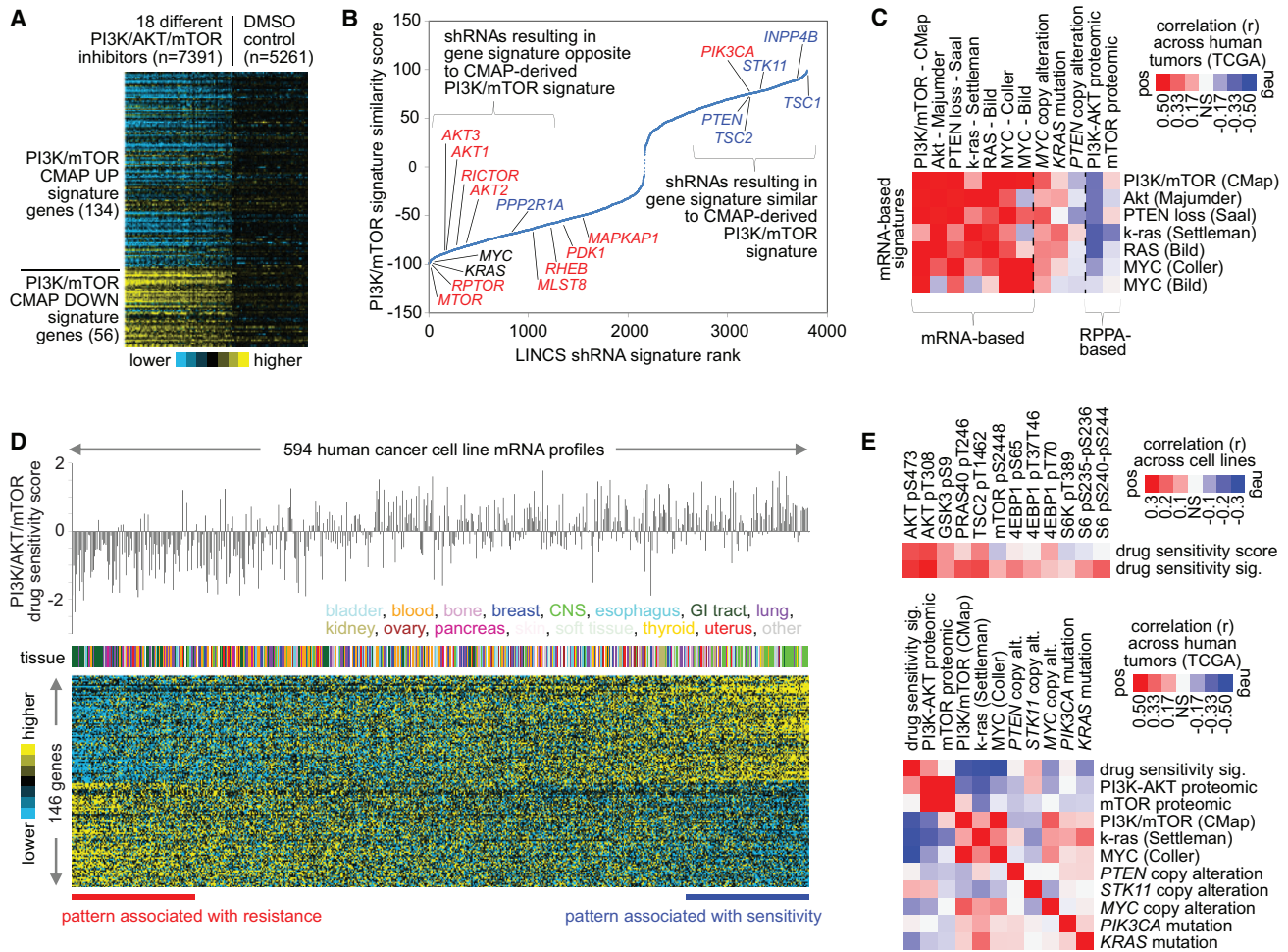


Figure 5. Survey of Two Distinct PI3K/AKT/mTOR-Associated Gene Transcription Signatures across Human Cancers

(A) A previously defined gene transcription signature of PI3K/AKT/mTOR (Creighton et al., 2010) (originally derived using the Connectivity Map, or CMAP, dataset) was re-examined in the LINCSeq database of perturbational expression profiles, with the PI3K/AKT/mTOR inhibitor treatment group compared with control group. (B) The PI3K/AKT/mTOR “CMAP” signature was evaluated against the LINCSeq expression profiles of cells treated with shRNAs for ~6,000 different genes. In the plot shown, shRNAs are ranked according to the overall similarities in their induced expression patterns with those of the PI3K/AKT/mTOR signature; for example, for shRNAs represented on the left of x axis, knock down of the gene results in a pattern inverse of that of the PI3K/AKT/mTOR signature. Red, canonical promoter of PI3K/AKT/mTOR pathway; blue, canonical suppressor. (C) TCGA pan-cancer mRNA profiles ($n = 10,224$ cases) were each scored for various transcriptional signatures associated with PI3K/AKT/mTOR, MYC, or k-ras pathways (defined previously using experimental models). Pearson’s correlations between indicated transcriptional and proteomic signature scores across the pan-cancer profiles are indicated, along with correlations of the signatures with specific genomic alterations. (D) A gene expression signature of sensitivity to PI3K/AKT/mTOR inhibition in cancer cell lines, consisting of 146 genes ($p < 0.01$ by t test and $p < 0.01$ in regression model incorporating tumor type as a confounder), was derived using the dataset of Garnett et al. (2012). (E) Top: for cell lines with both RPPA and mRNA data ($n = 231$), Pearson’s correlations between key PI3K/AKT/mTOR proteins and PI3K/AKT/mTOR inhibition sensitivity, as defined by either drug treatment or gene signature from (D). Bottom: TCGA pan-cancer mRNA profiles were each scored for the drug sensitivity signature from (D); Pearson’s correlations across the pan-cancer profiles, involving transcriptional and proteomic signature scores and selected genomic features, are indicated. See also Figure S5 and Table S6.

Genetic/Genomic Alteration Classes in Relation to PI3K/AKT/mTOR Pathway Activation

We then sought to examine the effects on pathway activation of some key genomic events in the tumors in which they occurred (including mutations represented in Figure 2A and copy alterations involving *PIK3CA*, *PTEN*, and *STK11*). Of the 7,099 tumor cases examined (with both mutation and protein data), 4,468 (63%) harbored at least one nonsilent somatic mutation or copy alteration involving PI3K/AKT/mTOR pathway (Figures 7A

and S6A). Another set of 764 tumors showed high levels of phospho-AKT (>0.5 SD of pS473 from the median across samples) but without any of the genetic or genomic alterations associated with the above 4,468 tumors, and another set of 394 tumors showed low levels of phospho-AMPK (<-0.5 SD) without an associated genetic or genomic alteration. In comparison with a set of tumors that did not show pathway alteration at the DNA or protein level (an “unaligned” set, $n = 1,058$), mutation or copy alteration of individual PI3K/AKT/mTOR pathway

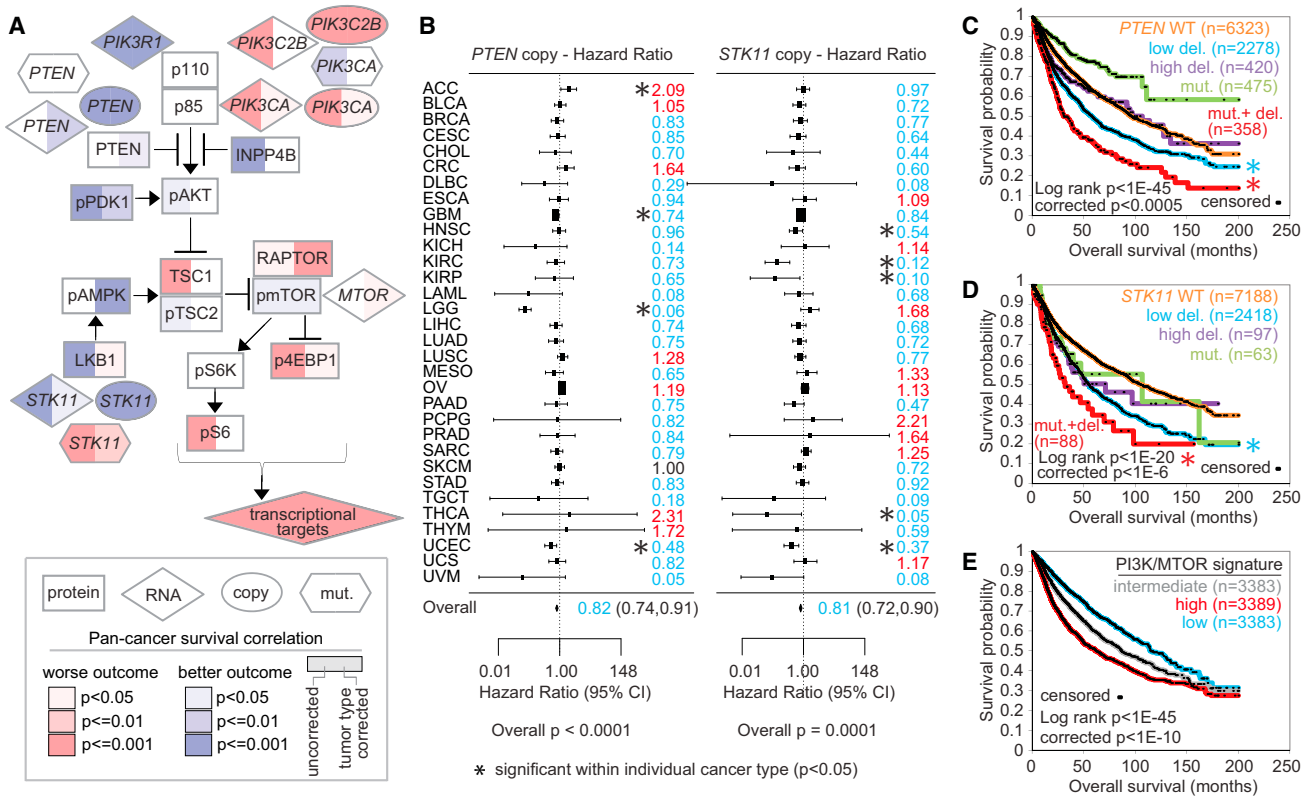


Figure 6. Pan-Cancer Molecular Correlates of Patient Survival Involving PI3K/AKT/mTOR Pathway Components

(A) Pathway diagram representing molecular features at the levels of mRNA (using $n = 10,152$ cancer cases in total with both mRNA and survival data), protein ($n = 7,532$), copy number ($n = 10,685$), and somatic mutation ($n = 10,054$). Red, significant correlation with worse patient outcome; blue, significant correlation with better outcome. “Tumor type corrected” survival p values denote significant correlation in model incorporating both the molecular feature and cancer type. p Values < 0.05 correspond to an estimated false discovery rate (Storey and Tibshirani, 2003) of $< 10\%$.

(B) Forest plots of hazard ratios by tumor type (with 95% confidence intervals) for patient death for *PTEN* copy alteration (left) and for *STK11* copy alteration (right). Hazard ratios based on log (tumor/normal) copy values; hazard ratio less than 1 (blue) denotes trend of copy loss with worse outcome. p Value for overall survival correlation by meta-analysis fixed effects model. Asterisks denote cancer types that were individually significant ($p < 0.05$).

(C and D) Kaplan-Meier plot of overall survival of patients stratified by *PTEN* (C) or *STK11* (D) alteration. Low del., low-level deletion (partial loss, no detected mutation); high del., high-level deletion (approximating total loss); mut., somatic nonsilent mutation (no copy loss); mut. + del., copy loss combined with mutation. Corrected p values by stratified log rank test incorporate cancer type as a confounder. Asterisks denote groups significantly different from wild-type (WT) group by stratified log rank test.

(E) Kaplan-Meier plot of overall survival of patients stratified by PI3K/AKT/mTOR transcriptional signature (CMAP signature). Corrected p values by stratified log rank test incorporate cancer type as a confounder.

members in general could be associated with higher PI3K/AKT or mTOR signaling as measured by protein arrays (Figures 7B and S6B). Notably, *STK11* alteration or low phospho-AMPK was strongly associated with high mTOR signaling, but not with high PI3K/AKT signaling, consistent with the LKB1/AMPK pathway acting on mTOR independently of PI3K/AKT (Figure 2A). Mutations associated with RTK signaling were not strongly associated with PI3K/AKT/mTOR activation (Figures 7A and 7B), indicative of decoupling between PI3K/AKT/mTOR and RTK. Low-level as well as high-level copy losses of *PTEN* and *STK11* could be associated with greater mTOR signaling.

PI3K/AKT/mTOR pathway activity, when measured at the protein level, was explained by known mutations or copy alteration in most but not all of the cases examined, suggesting additional, unexplained, or underappreciated mechanisms of pathway activation. Focusing on the “High P-AKT” tumor group ($n = 764$), with high phospho-AKT but lacking a DNA alteration classically

associated with PI3K/AKT activation, these tumors were highly enriched for specific cancer types including LGG, PRAD, KIRC, and PCPG (Figure 7C), as well as for *IDH1* mutations (associated primarily with LGG, i.e., gliomas) and *VHL* mutations (associated with renal cancers). A set of microRNAs could also help distinguish the “High P-AKT” group (Figure S6C). Proteins that were highly expressed specifically within the High P-AKT group (Figure 7D) included phospho-ERK, phospho-SRC, and phospho-NDRG1. These mutations and proteins would suggest a model (Figure 7E) whereby mutant *IDH1* may lead to high phospho-ERK (Chaturvedi et al., 2013) and SRC can activate PI3K (Chen et al., 2015; Su et al., 2016), and where activated mTOR signaling may activate transcription targets of hypoxia via HIF-1 α (particularly in the absence of VHL), including NDRG1 and growth factors that may lead to a further increase ERK and PI3K signaling (Clark, 2009). Notably, VHL was recently found to directly suppress AKT activity (Guo et al., 2016), and generation of

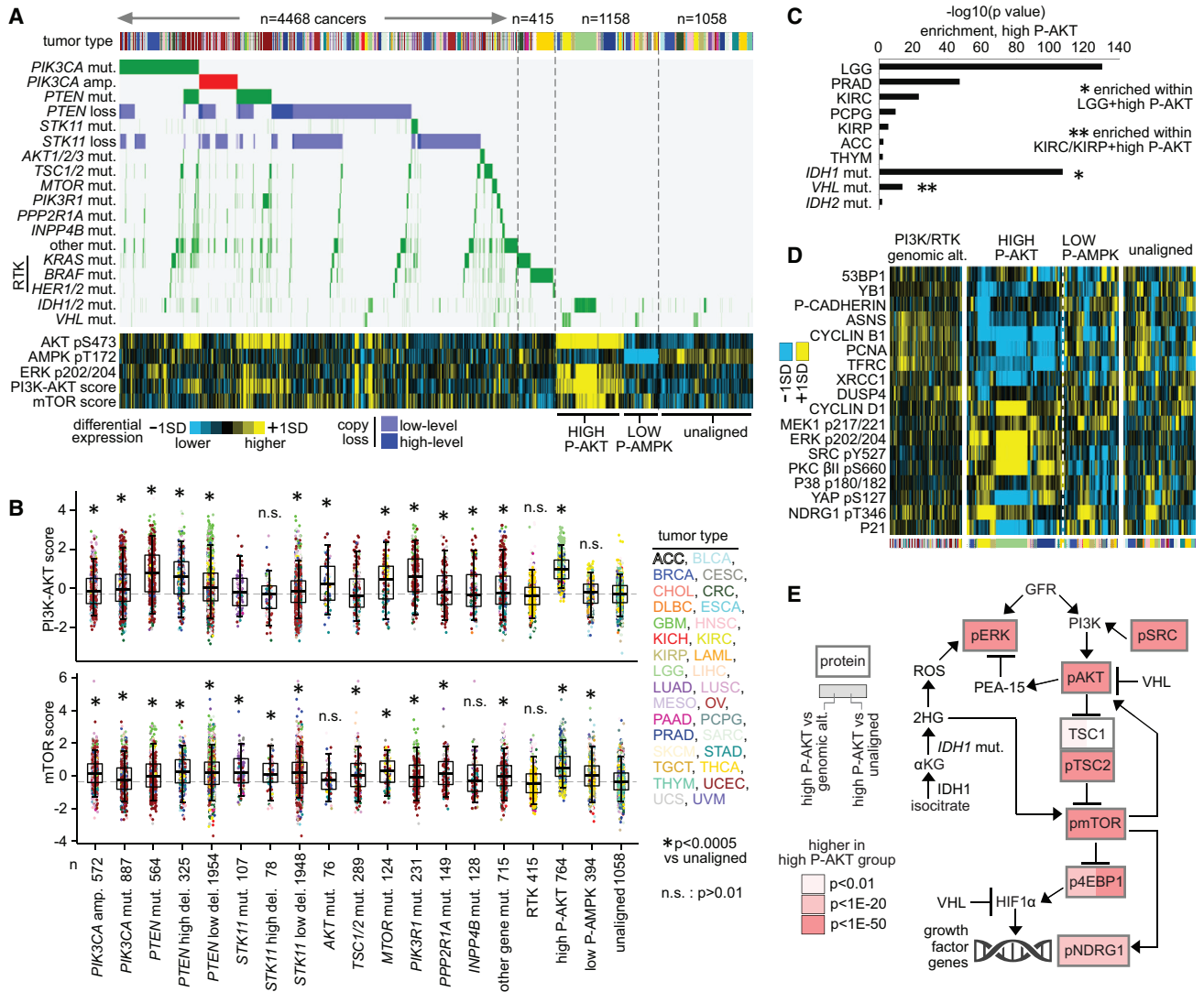


Figure 7. Tumor Classes as Defined by PI3K/AKT/mTOR-Related Alterations

(A) Tumor cases were separated into distinct groups on the basis of genetic or genomic alteration and of protein expression: (1) cases with nonsilent somatic mutation or copy alteration involving selected PI3K/AKT/mTOR pathway members as shown (left side, n = 4,468 cases), (2) additional cases with nonsilent mutation involving selected receptor tyrosine kinase (RTK)-associated genes (n = 415 cases), (3) cases with high phospho-AKT (HIGH P-AKT) but with none of the above somatic alterations (n = 764 cases), (4) cases with LOW phospho-AMPK (LOW P-AMPK) but with none of the above somatic alterations (n = 394 cases), (5) cases not aligned with any of the above (unaligned, n = 1,058 cases). AKT/MTOR/PIK3CA/PIK3R1/PTEN mutations represent “predicted functional” mutations from Figure 3D. Other mut. track involves nonsilent mutations for other genes represented in Figure 2A (STAR Methods and Figure S6). Protein values and proteomic scores normalized to SDs from the median.

(B) Boxplots of PI3K/AKT (top) and mTOR (bottom) pathway activity scores by alteration class. p Values by t test on log-transformed values. n.s., not significant (p > 0.01). Boxplots represent 5%, 25%, 50%, 75%, and 95%.

(C) Enriched tumor types and mutations within the HIGH P-AKT group. p Values by one-sided Fisher’s exact test. *IDH1* and *VHL* mutation events were significant (p < 1E-10 and p < 0.01, respectively) when limiting the analysis to LGG and to KIRC/KIRP (renal) cases, respectively.

(D) Top differentially expressed proteins in HIGH P-AKT group compared to unaligned and PI3K-altered groups (see STAR Methods), not including core PI3K/AKT/mTOR members.

(E) Diagram of interactions involving PI3K/AKT/MTOR pathway represented by selected features from (C) and (D) (Carbonneau et al., 2016; Dodd et al., 2015; Guo et al., 2016; Weiler et al., 2014), with differential protein expression patterns represented, comparing tumors in HIGH P-AKT group with tumors harboring PI3K/RTK genomic alteration or with unaligned tumors. p Values by t test on log-transformed data. See also Figure S6.

2-hydroxyglutarate by mutated *IDH1/2* was also recently found to lead to the activation of mTOR (Carbonneau et al., 2016); our data here would highlight the importance of both of the above relationships in the setting of human cancer.

DISCUSSION

TCGA pan-cancer datasets have enabled us to examine human tumor correlations in the context of PI3K/AKT/mTOR to an extent

not previously possible. Our current model of the PI3K/AKT/mTOR pathway has developed over the course of numerous independent molecular biology studies, spanning decades of research. In large part, our understanding of the pathway members and interactions involved has been derived from experimental systems, including cell lines. While cell lines may uncover cause-and-effect relationships *in vitro*, the relevance of such relationships in the setting of human diseases such as cancer may not always be clear from these data alone. On the other hand, molecular data from human tumors provide correlative (although not necessarily causal) relationships that would have relevance to disease in the human setting. Most of the correlations observed in our study fit well with our understanding of PI3K/AKT/mTOR signaling, in particular the genetic or genomic alteration of specific genes having an impact on phospho-protein expression of key downstream intermediates. Genes or alteration classes that were previously underappreciated would also be found relevant in our study, including partial loss of *PTEN* or *STK11* (associated with both worse survival and increased mTOR signaling). Where gene mutation often inactivates one allele, loss of one allele by copy alteration, which is common across multiple cancer types for both *PTEN* and *STK11*, would presumably have the same impact on loss of gene function. *IDH1* and *VHL* mutations would also be implicated here with PI3K/AKT/mTOR, where such alterations were associated with particularly high AKT/mTOR signaling, and which genes might be put forth for consideration as part of the “canon” of what would be recognized to constitute the core standard model of the PI3K/AKT/mTOR pathway.

The multiplatform molecular datasets offered by TCGA allow for a more comprehensive view of the PI3K/AKT/mTOR pathway. Pathway alterations in cancer may be manifested at different levels of molecular complexity, from DNA to protein to transcriptional consequences. Integration with RPPA proteomic data allows us to assess the impact on pathway activation of mutations or copy alterations observed at the DNA level. As observed in this study, multiple oncogenic pathways in addition to PI3K/AKT/mTOR may regulate similar sets of transcriptional targets, where transcriptional patterns would represent a degree of separation from the pathway as manifested at the protein level. Phospho-protein levels may only be assessed by protein data and not mRNA data, which also represents an advantage of RPPA compared with other proteomic approaches (Creighton and Huang, 2015). Clear overall trends may be observed when integrating proteomic data with data from other platforms, although statistical trends (e.g., visualized as boxplots) would apply to groups of patients and not always to the individual patient, which has implications regarding personalized therapy. Various sources of biological noise, in addition to technical noise, may be present within human tumors, which give rise to variation in molecular signals. Widespread molecular aberrations involving numerous genes and pathways within a given tumor, clonal heterogeneity, microenvironmental influences, variable sample purity, and tissue-specific effects can all add noise to our ability to match protein signals with specific DNA alterations. The RPPA methodology may have limitations as well (e.g., antibody robustness, unknown history of sample material used to measure potentially labile phosphorylations, linearity of signal readout, etc.), and instances where proteomic signals would

seem disconnected from other molecular profile features of a particular tumor may be difficult to interpret. The power of large sample numbers and the opportunities for data integration offered by TCGA pan-cancer cohort can aid greatly in detecting robust patterns relevant to our understanding aspects of pathway deregulation.

Results of this study include a comprehensive and annotated catalog of PI3K/AKT/mTOR-associated variants across over 10,000 tumors, which may serve as an additional resource for assessing variants in the clinical setting. One of the challenges of applying personalized and precision medicine approaches to cancer therapy is the large number of gene alterations that may be found within a given patient’s tumor. Stratifying patients by mutation status, e.g., *PIK3CA* mutation, has been shown to increase response rates in clinical trials testing inhibitors to PI3K/AKT/mTOR pathway, although non-responders are still common (Ilagan and Manning, 2016). Not all genetic variants impacting a given gene would necessarily have a similar impact on its function, including a large fraction of observed *PIK3CA* variants. Oncogenic variants that are found to occur frequently or are associated with a significant pattern would seem likely to be functionally relevant. Other measures of predicting variant functionality include *in silico* structural predictions, *in vitro* functional assays, domain-specific expertise, and protein expression, all of which were explored to varying extents in the present study. In practice, multiple measures may be needed, as no single measure may capture all of the variants likely to be functional. In addition, the RPPA proteomic platform would have potential for clinical applications to personalized therapy (Creighton and Huang, 2015), and transcriptional signatures associated with inhibitor sensitivity in cell lines may be defined (Singh et al., 2009). The focused, comprehensive analysis on the PI3K/AKT/mTOR pathway here will serve as a valuable resource for understanding its deregulation in cancers and how to maximize its clinical utility.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Human Subjects
 - Cell Lines
- METHOD DETAILS
 - TCGA Patient Cohort
 - Datasets
 - Gene and Protein Signatures
 - *In Silico* Mutation Evaluation
 - Cell Line Viability Assays
 - Tumor Classes by Gene Alteration
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ccell.2017.04.013>.

AUTHOR CONTRIBUTIONS

Conceptualization, C.J.C., D.J.K., and G.B.M.; Methodology, C.J.C., D.J.K., G.B.M., Y.Z., M.K., K.L.S., J.N.W., P.J.P., R.K., H.L., R.A., Y.H.T., and P.K.-S.N.; Investigation, Y.Z., P.K.-S.N., M.K., F.C., Y.L., Y.H.T., G.d.V., C.S.S., T.K.C., M.I., C.V.W., T.F.W., E.P.H. A.K.G., G.B.M., and C.J.C.; Formal Analysis, Y.Z., F.C., Y.L., M.K., P.K.-S.N., C.J.C., and D.J.K.; Data Curation, C.J.C., G.B.M., H.L., R.A., K.J.J., A.H., A.P., C.A.B., E.L., H.S.M., J.T., J.Z., L.Y., S.S., S.L., X.R., X.S., H.S., J.S., L.J.L., R.X., L.C., A.P., P.J.P., R.K., and P.K.-S.N.; Visualization, C.J.C. and F.C.; Writing – Original Draft, C.J.C. and D.J.K.; Writing – Review & Editing, G.B.M., C.J.C., D.J.K., P.K.-S.N., G.d.V., C.S.S., T.K.C., M.I., C.V.W., J.N.W., E.P.H., A.K.G., and H.L.; Supervision, C.J.C., D.J.K., and G.B.M.

ACKNOWLEDGMENTS

This work was supported in part by Cancer Prevention and Research Institute of Texas (CPRIT) grant RP120713 C2 (C.J.C.), and by NIH grants 5P50CA098258 (G.B.M.), 5U01CA168394 (G.B.M.), U24CA143883 (G.B.M.), 2P50CA101942-11A1 (D.J.K.), 1R21CA191687 (D.J.K.), 5P01CA120964 (D.J.K.), CA175486 (H.L.), CA209851 (H.L. and G.B.M.), P30CA125123 (C.J.C.), and Cancer Center Support Grant (CCSG) P30CA016672. G.B.M. has licensed an HRD assay to Myriad Genetics and is an SAB member/consultant with AstraZeneca, Catena Pharmaceuticals, Critical Outcome Technologies, ImmunoMET, Ionis, Medimmune, Nuevolution, Pfizer, Precision Medicine, Signalchem Lifesciences, Symphogen, Takeda/Millennium Pharmaceuticals, and Tarveda, and has stock options with Catena Pharmaceuticals, ImmunoMet, Spindle Top Ventures, and Tarveda. S.W. is a consultant with AstraZeneca, Clovis Oncology, Genentech, Medivation, Casdin Capital, and Vermillion. G.B.M. receives research support from Abbvie, Adelson Medical Research Foundation, AstraZeneca, the Breast Cancer Research Foundation, Critical Outcomes Technology, Horizon Diagnostics, Illumina, Ionis, Karus Therapeutics, Komen Research Foundation, Nanostring, Pfizer, Takeda/Millennium Pharmaceuticals, and Tesaro.

Received: January 10, 2017

Revised: March 17, 2017

Accepted: April 18, 2017

Published: May 18, 2017

REFERENCES

- Akbani, R., Ng, P., Werner, H., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J., Yoshihara, K., Li, J., et al. (2014). A pan-cancer proteomic perspective on the Cancer Genome Atlas. *Nat. Commun.* **5**, 3887.
- Carbonneau, M., Gagné, L., Lalonde, M., Germain, M., Motorina, A., Guiot, M., Secco, B., Vincent, E., Tumber, A., Hulea, L., et al. (2016). The oncometabolite 2-hydroxyglutarate activates the mTOR signalling pathway. *Nat. Commun.* **7**, 12700.
- Carpten, J.D., Faber, A.L., Horn, C., Donoho, G.P., Briggs, S.L., Robbins, C.M., Hostetter, G., Boguslawski, S., Moses, T.Y., Savage, S., et al. (2007). A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448**, 439–444.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B., Sumer, S., Aksoy, B., Jacobsen, A., Byrne, C., Heuer, M., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404.
- Chang, M., Asthana, S., Gao, S., Lee, B., Chapman, J., Kandath, C., Gao, J., Succi, N., Solit, D., Olshen, A., et al. (2016). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* **34**, 155–163.
- Chaturvedi, A., Araujo Cruz, M., Jyotsana, N., Sharma, A., Yun, H., Görlich, K., Wichmann, M., Schwarzer, A., Preller, M., Thol, F., et al. (2013). Mutant IDH1 promotes leukemogenesis in vivo and can be specifically targeted in human AML. *Blood* **122**, 2877–2887.
- Chen, B., Xu, X., Luo, J., Wang, H., and Zhou, S. (2015). Rapamycin enhances the anti-cancer effect of dasatinib by suppressing Src/PI3K/mTOR pathway in NSCLC Cells. *PLoS One* **10**, e0129663.
- Cheung, L.W., Hennessy, B.T., Li, J., Yu, S., Myers, A.P., Djordjevic, B., Lu, Y., Stemke-Hale, K., Dyer, M.D., Zhang, F., et al. (2011). High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* **1**, 170–185.
- Clark, P.E. (2009). The role of VHL in clear-cell renal cell carcinoma and its relation to targeted therapy. *Kidney Int.* **76**, 939–945.
- Creighton, C.J. (2008). Multiple oncogenic pathway signatures show coordinate expression patterns in human prostate tumors. *PLoS One* **3**, e1816.
- Creighton, C.J., and Huang, S. (2015). Reverse phase protein arrays in signaling pathways: a data integration perspective. *Drug Des. Dev. Ther.* **9**, 3519–3527.
- Creighton, C., Fu, X., Hennessy, B., Casa, A., Zhang, Y., Gonzalez-Angulo, A., Luch, A., Gray, J., Brown, P., Hilsenbeck, S., et al. (2010). Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast Cancer Res.* **12**, R40.
- Dibble, C.C., and Manning, B.D. (2013). Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. *Nat. Cell Biol.* **15**, 555–564.
- Dodd, K., Yang, J., Shen, M., Sampson, J., and Tee, A. (2015). mTORC1 drives HIF-1 α and VEGF-A signalling via multiple mechanisms involving 4E-BP1, S6K1 and STAT3. *Oncogene* **34**, 2239–2250.
- Dogruluk, T., Tsang, Y., Espitia, M., Chen, F., Chen, T., Chong, Z., Appadurai, V., Dogrulak, A., Eterovic, A., Bonnen, P., et al. (2015). Identification of variant-specific functions of PIK3CA by rapid phenotyping of rare mutations. *Cancer Res.* **75**, 5341–5354.
- Duan, Q., Flynn, C., Niepel, M., Hafner, M., Muhlich, J., Fernandez, N., Rouillard, A., Tan, C., Chen, E., Golub, T., et al. (2014). LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* **42**, W449–W460.
- Engelman, J.A., Luo, J., and Cantley, L.C. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat. Rev. Genet.* **7**, 606–619.
- Garnett, M., Edelman, E., Heidorn, S., Greenman, C., Dastur, A., Lau, K., Greninger, P., Thompson, I., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575.
- Grabner, B.C., Nardi, V., Birsoy, K., Possemato, R., Shen, K., Sinha, S., Jordan, A., Beck, A.H., and Sabatini, D.M. (2014). A diverse array of cancer-associated MTOR mutations are hyperactivating and can predict rapamycin sensitivity. *Cancer Discov.* **4**, 554–563.
- Guo, J., Chakraborty, A., Liu, P., Gan, W., Zheng, X., Inuzuka, H., Wang, B., Zhang, J., Zhang, L., Yuan, M., et al. (2016). pVHL suppresses kinase activity of Akt in a proline-hydroxylation-dependent manner. *Science* **353**, 929–932.
- Hanahan, D., and Weinberg, R. (2000). The hallmarks of cancer. *Cell* **100**, 57–70.
- Hennessy, B., Smith, D., Ram, P., Lu, Y., and Mills, G. (2005). Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nat. Rev. Drug Discov.* **4**, 988–1004.
- Hoadley, K., Yau, C., Wolf, D., Cherniack, A., Tamborero, D., Ng, S., Leiserson, M., Niu, B., McLellan, M., Uzunangelov, V., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944.
- Hornigold, N., Devlin, J., Davies, A.M., Aveyard, J.S., Habuchi, T., and Knowles, M.A. (1999). Mutation of the 9q34 gene TSC1 in sporadic bladder cancer. *Oncogene* **18**, 2657–2661.
- Ilagan, E., and Manning, B. (2016). Emerging role of mTOR in the response to cancer therapeutics. *Trends Cancer* **2**, 241–251.
- Kandath, C., McLellan, M., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J., Wyczalkowski, M., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339.

- Keniry, M., and Parsons, R. (2008). The role of PTEN signaling perturbations in cancer and in targeted therapy. *Oncogene* *27*, 5477–5485.
- Laplante, M., and Sabatini, D.M. (2012). mTOR signaling in growth control and disease. *Cell* *149*, 274–293.
- Lawrence, M., Stojanov, P., Mermel, C., Robinson, J., Garraway, L., Golub, T., Meyerson, M., Gabriel, S., Lander, E., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* *505*, 495–501.
- Li, J., Zhao, W., Akbani, R., Liu, W., Ju, Z., Ling, S., Vellano, C., Roebuck, P., Yu, Q., Eterovic, A., et al. (2017). Characterization of human cancer cell lines by reverse-phase protein arrays. *Cancer Cell* *31*, 225–239.
- Manning, B.D., and Cantley, L.C. (2007). AKT/PKB signaling: navigating downstream. *Cell* *129*, 1261–1274.
- Mayer, I.A., and Arteaga, C.L. (2016). The PI3K/AKT pathway as a target for cancer treatment. *Annu. Rev. Med.* *67*, 11–28.
- Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* *39*, e118.
- Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S.M., Riggins, G.J., et al. (2004). High frequency of mutations of the PIK3CA gene in human cancers. *Science* *304*, 554.
- Singh, A., Greninger, P., Rhodes, D., Koopman, L., Violette, S., Bardeesy, N., and Settleman, J. (2009). A gene expression signature associated with “K-Ras addiction” reveals regulators of EMT and tumor cell survival. *Cancer Cell* *15*, 489–500.
- Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* *100*, 9440–9445.
- Su, W., Chen, Y., Wang, C., Ding, X., Rwibasira, G., and Kong, Y. (2016). Human cathelicidin LL-37 inhibits platelet aggregation and thrombosis via Src/PI3K/Akt signaling. *Biochem. Biophys. Res. Commun.* *473*, 283–289.
- The_Cancer_Genome_Atlas_Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* *490*, 61–70.
- The_Cancer_Genome_Atlas_Research_Network. (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* *499*, 43–49.
- Thorpe, L.M., Yuzugullu, H., and Zhao, J.J. (2015). PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat. Rev. Cancer* *15*, 7–24.
- Weiler, M., Blaes, J., Pusch, S., Sahm, F., Czabanka, M., Luger, S., Bunse, L., Solecki, G., Eichwald, V., Jugold, M., et al. (2014). mTOR target NDRG1 confers MGMT-dependent resistance to alkylating chemotherapy. *Proc. Natl. Acad. Sci. USA* *111*, 409–414.
- Yang, L., Luquette, L., Gehlenborg, N., Xi, R., Haseley, P., Hsieh, C., Zhang, C., Ren, X., Protopopov, A., Chin, L., et al. (2013). Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* *153*, 919–929.
- Zack, T., Schumacher, S., Carter, S., Cherniack, A., Saksena, G., Tabak, B., Lawrence, M., Zhsng, C., Wala, J., Mermel, C., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* *45*, 1134–1140.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
TCGA whole exome DNA sequence data	Unified Ensemble “MC3” Call Set from DNA Nexus	https://www.synapse.org/#!Synapse:syn7214402
TCGA protein expression data by RPPA	TCPA Portal, Level 4	http://tcpaportal.org/tcpa/
TCGA RNA expression by RNA-seq	Broad Firehose Datasets	gdac.broadinstitute.org
DNA copy number alteration by Affymetrix SNP 6 array	Broad Firehose Datasets	https://gdac.broadinstitute.org
TCGA whole genome DNA sequence	Genomic Data Commons	https://gdc.cancer.gov/
Perturbational expression profiles (compounds, shRNAs)	BROAD LINCS database	http://www.lincsproject.org/
mRNA and drug sensitivity measurements in cancer cell lines	Genomics of Drug Sensitivity in Cancer Portal	http://www.cancerrxgene.org/
RPPA profiles of cancer cell lines	MCLP Data Portal	http://tcpaportal.org/mclp/
Experimental Models: Cell Lines		
MCF10A	ATCC	Authenticated by Short Tandem Repeat (STR) analysis at M.D. Anderson Characterized Cell Line Core facility (Houston, TX)
Ba/F3	M.D. Anderson Characterized Cell Line Core facility (Houston, TX)	Parental cells validated based on continued dependence on IL3 for propagation (mouse-originated cell line)

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Chad J. Creighton (creight@bcm.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

Cancer molecular profiling data were generated through informed consent as part of previously published studies and analyzed in accordance with each original study’s data use guidelines and restrictions.

Cell Lines

Assay medium for survival assay were Advanced RPMI 1640 medium (Life Technologies) with 5% FBS (Life Technologies) and 1x GlutaMAX (Life Technologies) for Ba/F3 cells and MEBM Basal medium (Lonza) with 100 ng/ml Cholera toxin (Lonza) and 52 ng/ml Bovine Pituitary Extract (BPE) (Lonza) for MCF10A cells.

METHOD DETAILS

TCGA Patient Cohort

The results here are based upon data generated by TCGA Research Network (<http://cancergenome.nih.gov/>). Molecular data from 11219 human cancers were aggregated from public repositories (Table S1). Tumors spanned 32 different TCGA projects, each project representing a specific cancer type, listed as follows: LAML, Acute Myeloid Leukemia; ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; LGG, Brain Lower Grade Glioma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; CRC, Colorectal adenocarcinoma (combining COAD and READ projects); ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; DLBC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD,

Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; PRAD, Prostate adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular Germ Cell Tumors; THYM, Thymoma; THCA, Thyroid carcinoma; UCS, Uterine Carcinosarcoma; UCEC, Uterine Corpus Endometrial Carcinoma.

Datasets

Proteomic data were generated by RPPA across 7663 patient tumors obtained from TCGA. RPPA methodology and quality control procedures have been described previously (Akbani et al., 2014; Li et al., 2017). In total, 225 high-quality antibodies targeting total (n=166), cleaved (n=2), acetylated (n=1) and phosphoproteins (n=56) were used. The entire set of RPPA Pan-Cancer samples was run in several different batches, resulting in potential batch effects on merging the sets; replicates-based normalization (RBN) (Akbani et al., 2014), was therefore applied, using replicate samples run across multiple batches to adjust the data for batch effects. Data (“Level 4”) are available from The Cancer Proteome Atlas (<http://tcpportal.org/tcpa>).

RNA-seq and miRNA-seq data were obtained from The Broad Institute Firehose pipeline (<http://gdac.broadinstitute.org>). All RNA-seq samples were aligned using the by UNC RNA-seq V2 pipeline (The Cancer Genome Atlas Research Network, 2013). For miRNA-seq data, only sample profiles from the Hiseq platform were used (representing n=8690 cases).

DNA from each tumor or germline-derived sample was hybridized to Affymetrix SNP 6.0 arrays as previously described (The Cancer Genome Atlas Research Network, 2013) (n=10845 tumor profiles in all). GISTIC 2.0 was applied to the transformed copy number data, with a noise threshold used to determine copy gain or loss. Low-level gene gain, high-level gene amplification, low-level copy loss, or high-level copy loss were inferred using the “thresholded” calls as made by Broad Firehose pipeline (using +1, +2, -1, or -2, respectively). High-level amplifications denotes amplifications above the threshold and larger than the arm level amplifications observed for the given sample. Low-level copy deletions represent deletion above the threshold (approximating heterozygous deletions in the absence of whole genome doubling); high-level copy deletions denote copy losses above the threshold and greater than the minimum arm-level deletion observed for the sample (approximating homozygous deletions in the absence of whole genome doubling). Log (tumor/normal) copy values were used to evaluate correlations with survival in Figure 6A.

Somatic mutation calls were obtained from the publicly-available “MC3” TCGA MAF file (covering n=10224 patients, <https://www.synapse.org/#!Synapse:syn7214402>). This MC3 set is a re-calling of uniform files from all TCGA projects, with variant calling using a standardized set of mutation callers. The BAM files used underwent a standardized local re-alignment to hg19 (Genome Reference Consortium GRCh37), six calling algorithms were applied, and a number of automated filters were applied. Variants called by two or more algorithms were used in the study. Whole genome sequence analysis was carried out for 1363 cases (with paired normal samples, high pass coverage for BRCA and OV cases, low pass for BLCA, CESC, CRC, ESCA, HNSC, LGG, LUAD, PRAD, SKCM, STAD, THCA, UCEC, and UVM). Genomic rearrangements were detected in all tumor and normal genomes by Meerkat (Yang et al., 2013). Five discordant read pairs support are required for each event. Each event was detected in tumor genome was filtered by all normal genomes to ensure it represented a somatic event.

Gene and Protein Signatures

Pan-cancer RPPA profiles were scored for a PI3K/AKT pathway signature, defined as the sum of normalized phosphoprotein levels of AKT (both S473 and T308 RPPA features), GSK3 (S9 and S21/S9 features), PRAS40, and phospho-TSC2. RPPA profiles were also scored for an mTOR pathway signature, defined as the sum of phosphoprotein levels of mTOR, 4EBP1 (S65, T37/T46, and T70 RPPA features), P70S6K, and S6 (S235/S236 and S240/S244 features).

Gene transcriptional signatures of PI3K/AKT/mTOR pathway were defined as described previously (Creighton et al., 2010): “Saal” PTEN loss signature, genes correlated with Pten protein levels in breast cancer; “CMap” PI3K/AKT/mTOR signature, genes modulated in vitro by inhibitors to PI3K or mTOR, according to CMap dataset ($p < 0.01$, comparing PI3K/mTOR-inhibited cells with the rest of the Cmap profiles); “Majumder” Akt signature, genes modulated in a mouse model of inducible AKT ($p < 0.01$). MYC signatures (Coller and Bild) and the Bild Ras signature were from ref (Creighton, 2008), and the Settleman k-ras sensitivity signature were from ref (Singh et al., 2009). For a given gene transcription signature, we extracted the expression values from the TCGA gene expression array dataset. For each gene, we normalized expression values to standard deviations from the median across tumors. For signatures with “up” versus “down” genes, we computed our previously described “t-score” (Creighton et al., 2010) to score each tumor profile for relative manifestation of the signature.

For deriving a PI3K/AKT/mTOR drug sensitivity signature in cell lines (Figure 5D), we utilized the dataset from Garnett et al. (Garnett et al., 2012). For the 11 inhibitors to PI3K/AKT/mTOR represented in Garnett (including Rapamycin:MTOR, JW-7-52-1:MTOR, A-443654:AKT1/2/3, CHIR-99021:GSK3B, AZD6482:PI3Kb (P3C2B), AKT inhibitor VIII:AKT1/2, Temsirolimus:MTOR, MK-2206:AKT1/2, NVP-BE2235:PI3K (class 1) and mTORC1/2, GDC0941:PI3K (class 1), and AZD8055:mTORC1/2), we normalized IC_{50} values to standard deviations from median, then average to get single drug sensitivity score. Each gene was correlated in expression with the drug sensitivity score, first selecting for genes significant with $p < 0.01$ by t-test on log-transformed data (1099 significant genes), then further selecting for genes remaining significant after correcting for tissue type differences using a regression model that incorporates tumor type as a confounder (146 genes with corrected $p < 0.01$).

In Silico Mutation Evaluation

In assessing whether mutations may be more or less likely to have a functional effect on the resulting protein, a number of factors were considered. Somatic substitution hotspots (470 in total involving 275 genes), based on a previous pan-cancer analysis of

11119 human tumors (Chang et al., 2016), were incorporated into the present study where noted. Mutation Assessor calls predicting the functional impact (medium to high) of somatic mutation (Reva et al., 2011) were obtained from cBioPortal (Cerami et al., 2012). Manual review of variants involving *AKT1/2/3*, *MTOR*, *PIK3CA*, *PTEN*, *RHEB*, *TSC1/2* was also carried out by domain experts in the analysis group. Mutations that were predicted as potentially functional by any of the above—as well as mutations in tumor suppressor genes (e.g. *PTEN*, *PIK3R1*) classified as nonsense, frameshift, or indel—were evaluated separately with respect to comparing with AKT pS473 phospho-protein expression (Figure 3D).

Cell Line Viability Assays

The effects of mutations on the function of PIK3CA and PIK3R1 were assessed in Ba/F3 and MCF10A by survival assay as previously described (Dogruluk et al., 2015) with lentiviral vector pHAGE used in the cloning. In Ba/F3, the PIK3CA mutations were assigned as “Strong activating (SA)” if the mutations have an activity higher than M1043I (known moderate driver); as “Moderate activating (MA)” if the mutations have a similar or lower activity than M1043I; as “No difference from WT (NDFW)” if the mutations have a similar activity with WT; or as “Inactivating (INA)” if the mutations have an activity similar to negative controls (GFP/mCherry/Luciferase). The PIK3R1 mutations were assigned as “SA” if the mutations have a relative level of activation higher than that of PIK3CA M1043I comparing to negative controls; as “MA” if the mutations have a relative level of activation between PIK3CA M1043I and WT; as “Weak activating (WA)” if the mutations have a relative level of activation between PIK3CA WT and negative controls; or as “NDFW” if the mutations have a similar activity with WT. In MCF10A, the PIK3CA mutations were assigned as “SA” and “NDFW” by the same mean as in Ba/F3 model. The mutations were assigned as “MA” and “WA” if the mutations have an activity above and lower than 50% of that of M1043I, respectively.

Tumor Classes by Gene Alteration

Genetic/genomic alteration classes in relation to PI3K/AKT/mTOR pathway alteration were defined (Figure 7A), in order to relate these to PI3K/AKT and mTOR activation, as defined by protein signature score. For AKT/MTOR/PIK3CA/PIK3R1/PTEN mutations, “predicted functional” mutations from Figure 3D were used. An “other gene mutation” class of Figures 7A and 7B involved nonsilent mutations for other genes represented in Figure 2A (*AKT5*, *DEPDC5*, *DEPTOR*, *MAPKAP1*, *MLST8*, *NPRL2*, *NPRL3*, *PDK1*, *PRR5*, *RHEB*, *RICTOR*, *RPTOR*, *PIK3C2B*). The RTK group represented cases with hotspot mutations in *KRAS*, *BRAF*, *EGFR*, or *ERBB2*, that were not also included in the other PI3K/AKT/mTOR-related groups. The set of genes previously found significantly mutated in pan-cancer analysis (Lawrence et al., 2014), were searched for enrichment of mutation events within the High P-AKT group (Figure 7C). When defining proteins that were highly expressed specifically within the High P-AKT group (Figure 7D), RPPA features were selected that were over- or under-expressed in the High P-AKT compared to unaligned cases ($p < 0.05$, t-test on log-transformed data) for at least four of the seven cancer types, and differentially expressed in High P-AKT compared to unaligned and to PI3K/AKT/mTOR or RTK-altered cases across all cancer cases ($p < 0.01$ for each).

QUANTIFICATION AND STATISTICAL ANALYSIS

All p values were two-sided unless otherwise specified. Statistical significance was defined at the 0.05 threshold. All available TCGA data in the public domain at the time of this study was utilized, and no patients were deliberately excluded. Differential expression between comparison groups was assessed using t-test on log-transformed values. For visualization using heat maps and box plots, mRNA and protein expression values were z-normalized to standard deviations from the median across all tumor sample profiles.

Individual gene and protein features were evaluated for correlation with patient survival by univariate Cox analysis; in addition, a stratified Cox model was used to evaluate survival association when correcting for tumor type. For *PTEN* and *STK11* copy alteration features (Log [tumor/normal] ratios), Cox regression analysis within each individual cancer type was carried out; then, in order to aggregate the results across cancer types, we used “metafor” R package to conduct meta-analyses, with a random-effects model used to estimate the overall effectiveness of the molecular feature. For Kaplan-Meier plots, a stratified Log-rank test evaluated differences between tumor groups after correction for tumor type. Patient survival data from TCGA were current as of March 31, 2016.