Manuscript Title

This manuscript (permalink) was automatically generated from greenelab/mpmp-manuscript@296df81 on April 7, 2021.

Authors

- **John Doe**
  XXXX-XXXX-XXXX-XXXX · johndoe · johndoe
  Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe**
  XXXX-XXXX-XXXX-XXXX · janeroe
  Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

Introduction

Although cancer can be initiated and driven by many different genetic alterations, these tend to converge on a limited number of pathways or signaling processes [1]. A comprehensive understanding of how diverse genetic alterations perturb these central pathways is vital to precision medicine and biomarker identification efforts, as driver mutation status alone confers limited prognostic information [2,3]. While many methods exist to distinguish driver mutations from passenger mutations based on genomic sequence characteristics [4,5,6], until recently it has been a challenge to connect driver mutations to downstream changes in gene expression and cellular function within individual tumor samples.

The Cancer Genome Atlas (TCGA) Pan-Cancer Atlas provides uniformly processed, multi-platform -omics measurements across tens of thousands of samples from 33 cancer types [7]. Enabled by this publicly available data, a growing body of work on linking the presence of driving genetic alterations in cancer to downstream gene expression changes has emerged. Recent studies have considered Ras pathway alteration status in colorectal cancer [8], alteration status across many cancer types in Ras genes [9,10], TP53 [11], and PIK3CA [12], and alteration status across cancer types in frequently mutated genes [13]. More broadly, other groups have drawn on similar ideas to distinguish between the functional effects of different alterations in the same driver gene [14], to link alterations with similar gene expression signatures within cancer types [15], and to identify trans-acting expression quantitative trait loci (trans-eQTLs) in germline genetic studies [16].

These studies share a common thread: they each combine genomic (point mutation and copy number variation) data with transcriptomic (RNA sequencing) data within samples to interrogate the functional effects of genetic variation. RNA sequencing is ubiquitous and cheap, and its experimental and computational methods are relatively mature, making it a vital tool for generating insight into cancer pathology [17]. Some driver mutations, however, are known to act indirectly on gene expression through varying mechanisms. For example, oncogenic IDH1 and IDH2 mutations in glioma have been shown to interfere with histone demethylation, which results in increased DNA methylation and blocked cell differentiation [18,19]. Other genes implicated in aberrant DNA methylation in cancer include the TET family of genes [20] and SETD2 [21]. Certain driver mutations, such as those in DNA damage repair genes, may lead to detectable patterns of somatic mutation [22]. Additionally, correlation between gene expression and protein abundance in cancer cell lines is limited, and proteomics data could correspond more directly to certain cancer phenotypes and pathway perturbations [23]. In these contexts and others, integrating different data modalities or combining multiple data modalities could be more effective than relying solely on gene expression as a functional signature.

Here, we seek to compare -omics data types profiled in the TCGA Pan-Cancer Atlas for use as a multivariate functional readout of genetic alterations in cancer. We focus on DNA methylation (27K and 450K probe chips), reverse phase protein array (RPPA), and mutational signatures data [24] as alternative readouts. Prior studies have identified univariate correlations of CpG site methylation [25,26] and correlations of RPPA protein profiles [27] with the presence or absence of certain driver mutations. Other relevant past work includes linking point mutations and copy number variants (CNVs) with changes in methylation and expression at individual genes [28,29] and identifying functional modules that are perturbed by somatic mutations [30,31]. However, no direct comparison has been made between different data types for this application, particularly in the multivariate case where we consider changes to -omics-derived gene signatures rather than individual genes in isolation.

We select a wide-ranging collection of potential cancer drivers with varying functions and roles in cancer development [32]. We use mutation status in these genes as labels to train classifiers, using each of the data types listed as training data, in a pan-cancer setting; we follow similar methods to the elastic net logistic regression approach described in Way et al. 2018 [9] and Way et al. 2020 [13]. We show that although there is considerable predictive signal for many genes in each dataset relative to a random baseline, gene expression data tends to provide more effective predictions than the other data types in the vast majority of cases. In addition, we observe that combining data types into a single multi-omics model provides little, if any, performance benefit over the most performant model using a single data type. Our results will help to inform the design of future functional genomics studies in cancer, suggesting that RNA sequencing can serve as a broadly effective first-line readout for a variety of genetic alterations.

References

1. **Oncogenic Signaling Pathways in The Cancer Genome Atlas**
   Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K. Chatila, Augustin Luna, Konnor C. La, Sofia Dimitriadoy, David L. Liu, Havish S. Kantheti, Sadegh Saghafinia, … Armaz Mariamidze
   Cell (2018-04) https://doi.org/gc7r9b
   DOI: 10.1016/j.cell.2018.03.035 · PMID: 29625050 · PMCID: PMC6070353

2. **Systematic identification of mutations and copy number alterations associated with cancer patient prognosis**
   Joan C Smith, Jason M Sheltzer
   eLife (2018-12-11) https://doi.org/gf4zgg
   DOI: 10.7554/elife.39217 · PMID: 30526857 · PMCID: PMC6289580

3. **Challenges in identifying cancer genes by analysis of exome sequencing data**
   Matan Hofree, Hannah Carter, Jason F. Kreisberg, Sourav Bandyopadhyay, Paul S. Mischel, Stephen Friend, Trey Ideker
   Nature Communications (2016-07-15) https://doi.org/f8x7t3
   DOI: 10.1038/ncomms12096 · PMID: 27417679 · PMCID: PMC4947162

4. **Evaluating the evaluation of cancer driver genes**
   Collin J. Tokheim, Nickolas Papadopoulos, Kenneth W. Kinzler, Bert Vogelstein, Rachel Karchin
   Proceedings of the National Academy of Sciences (2016-12-13) https://doi.org/f9d77w
   DOI: 10.1073/pnas.1616440113 · PMID: 27911828 · PMCID: PMC5167163

5. **Detailed modeling of positive selection improves detection of cancer driver genes**
   Siming Zhao, Jun Liu, Pranav Nanga, Yuwen Liu, A. Ercument Cicek, Nicholas Knoblauch, Chuan He, Matthew Stephens, Xin He
   Nature Communications (2019-07-30) https://doi.org/gjmhnn
   DOI: 10.1038/s41467-019-11284-9 · PMID: 31363082 · PMCID: PMC6667447

6. **Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers**
   Ruth Nussinov, Hyunbum Jang, Chung-Jung Tsai, Feixiong Cheng
   PLOS Computational Biology (2019-03-28) https://doi.org/gg8jhm
   DOI: 10.1371/journal.pcbi.1006658 · PMID: 30921324 · PMCID: PMC6438456

7. **The Cancer Genome Atlas Pan-Cancer analysis project**
   John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, The Cancer Genome Atlas Research Network
   Nature Genetics (2013-09-26) https://doi.org/f3nt5c
   DOI: 10.1038/ng.2764 · PMID: 24071849 · PMCID: PMC3919969

8. **Modeling RAS Phenotype in Colorectal Cancer Uncovers Novel Molecular Traits of RAS Dependency and Improves Prediction of Response to Targeted Agents in Patients**
   Justin Guinney, Charles Ferté, Jonathan Dry, Robert McEwen, Gilles Manceau, KJ Kao, Kai-Ming Chang, Claus Bendtsen, Kevin Hudson, Erich Huang, … Pierre Laurent-Puig
   Clinical Cancer Research (2014-01-01) https://doi.org/f5njhn
   DOI: 10.1158/1078-0432.ccr-13-1943 · PMID: 24170544 · PMCID: PMC4141655

9. **Machine Learning Detects Pan-cancer Ras Pathway Activation in The Cancer Genome Atlas**
Gregory P. Way, Francisco Sanchez-Vega, Konnor La, Joshua Armenia, Walid K. Chatila, Augustin Luna, Chris Sander, Andrew D. Cherniack, Marco Mina, Giovanni Ciriello, … Armaz Mariamidze
Cell Reports (2018-04) https://doi.org/gfspsb
DOI: 10.1016/j.celrep.2018.03.046 · PMID: 29617658 · PMCID: PMC5918694

10. **Identification of pan-cancer Ras pathway activation with deep learning**
Xiangtao Li, Shaochuan Li, Yunhe Wang, Shixiong Zhang, Ka-Chun Wong
Briefings in Bioinformatics (2020-10-30) https://doi.org/gjmd3p
DOI: 10.1093/bib/bbaa258 · PMID: 33126245

11. **Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas**
Theo A. Knijnenburg, Linghua Wang, Michael T. Zimmermann, Nyasha Chambwe, Galen F. Gao, Andrew D. Cherniack, Huihui Fan, Hui Shen, Gregory P. Way, Casey S. Greene, … Armaz Mariamidze
Cell Reports (2018-04) https://doi.org/gfspsc
DOI: 10.1016/j.celrep.2018.03.076 · PMID: 29617664 · PMCID: PMC5961503

12. **Prediction of PIK3CA mutations from cancer gene expression data**
Jun Kang, Ahwon Lee, Youn Soo Lee
PLOS ONE (2020-11-09) https://doi.org/gjmd3s
DOI: 10.1371/journal.pone.0241514 · PMID: 33166334 · PMCID: PMC7652327

13. **Compressing gene expression data using multiple latent space dimensionalities learns complementary biological representations**
Gregory P. Way, Michael Zietz, Vincent Rubinetti, Daniel S. Himmelstein, Casey S. Greene
Genome Biology (2020-05-11) https://doi.org/gg2mjh
DOI: 10.1186/s13059-020-02021-3 · PMID: 32393369 · PMCID: PMC7212571

14. **Systematic interrogation of mutation groupings reveals divergent downstream expression programs within key cancer genes**
Michal R. Grzadkowski, Hannah Manning, Julia Somers, Emek Demir
Cold Spring Harbor Laboratory (2020-06-18) https://doi.org/gjmd7t
DOI: 10.1101/2020.06.02.128850

15. **Using Transcriptional Signatures to Find Cancer Drivers with LURE**
David Haan, Ruikang Tao, Verena Friedl, Ioannis N Anastopoulos, Christopher K Wong, Alana S Weinstein, Joshua M Stuart
World Scientific Pub Co Pte Lt (2019-12) https://doi.org/gjmd4t
DOI: 10.1142/9789811215636_0031

16. **Reverse regression increases power for detecting trans-eQTLs**
Saikat Banerjee, Franco L. Simonetti, Kira E. Detrois, Anubhav Kaphle, Raktim Mitra, Rahul Nagial, Johannes Söding
Cold Spring Harbor Laboratory (2020-09-02) https://doi.org/gjmhd8
DOI: 10.1101/2020.05.07.083386

17. **Cancer transcriptome profiling at the juncture of clinical translation**
Marcin Cieślik, Arul M. Chinnaiyan
Nature Reviews Genetics (2017-12-27) https://doi.org/gcsmnr
DOI: 10.1038/nrg.2017.96 · PMID: 29279605

18. ***IDH1* and *IDH2* Mutations in Gliomas**
Hai Yan, D. Williams Parsons, Genglin Jin, Roger McLendon, B. Ahmed Rasheed, Weishi Yuan, Ivan Kos, Ines Batinic-Haberle, Siân Jones, Gregory J. Riggins, … Darell D. Bigner
New England Journal of Medicine (2009-02-19) https://doi.org/btz6db
DOI: 10.1056/nejmoa0808710 · PMID: 19228619 · PMCID: PMC2820383

19. **IDH mutation impairs histone demethylation and results in a block to cell differentiation**
Chao Lu, Patrick S. Ward, Gurpreet S. Kapoor, Dan Rohle, Sevin Turcan, Omar Abdel-Wahab, Christopher R. Edwards, Raya Khanin, Maria E. Figueroa, Ari Melnick, … Craig B. Thompson
Nature (2012-02-15) https://doi.org/f4msnt
DOI: 10.1038/nature10860 · PMID: 22343901 · PMCID: PMC3478770

20. **Connections between TET proteins and aberrant DNA modification in cancer**
Yun Huang, Anjana Rao
Trends in Genetics (2014-10) https://doi.org/f6jm7v
DOI: 10.1016/j.tig.2014.07.005 · PMID: 25132561 · PMCID: PMC4337960

21. **SETting the Stage for Cancer Development: SETD2 and the Consequences of Lost Methylation**
Catherine C. Fahey, Ian J. Davis
Cold Spring Harbor Perspectives in Medicine (2017-05) https://doi.org/gjmfvq
DOI: 10.1101/cshperspect.a026468 · PMID: 28159833 · PMCID: PMC5411680

22. **Mechanisms underlying mutational signatures in human cancers**
Thomas Helleday, Saeed Eshtad, Serena Nik-Zainal
Nature Reviews Genetics (2014-07-01) https://doi.org/f25gnp
DOI: 10.1038/nrg3729 · PMID: 24981601 · PMCID: PMC6044419

23. **Quantitative Proteomics of the Cancer Cell Line Encyclopedia**
David P. Nusinow, John Szpyt, Mahmoud Ghandi, Christopher M. Rose, E. Robert McDonald, Marian Kalocsay, Judit Jané-Valbuena, Ellen Gelfand, Devin K. Schweppe, Mark Jedrychowski, … Steven P. Gygi
Cell (2020-01) https://doi.org/ggxbh5
DOI: 10.1016/j.cell.2019.12.023 · PMID: 31978347 · PMCID: PMC7339254

24. **The repertoire of mutational signatures in human cancer**
Ludmil B. Alexandrov, Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, Kyle R. Covington, Dmitry A. Gordenin, Erik N. Bergstrom, … PCAWG Consortium
Nature (2020-02-05) https://doi.org/ggkfnv
DOI: 10.1038/s41586-020-1943-3 · PMID: 32025018 · PMCID: PMC7054213

25. **Significant associations between driver gene mutations and DNA methylation alterations across many cancer types**
Yun-Ching Chen, Valer Gotea, Gennady Margolin, Laura Elnitski
PLOS Computational Biology (2017-11-10) https://doi.org/gchz8h
DOI: 10.1371/journal.pcbi.1005840 · PMID: 29125844 · PMCID: PMC5709060

26. **A pan-cancer analysis of driver gene mutations, DNA methylation and gene expressions reveals that chromatin remodeling is a major mechanism inducing global changes in cancer epigenomes**
Ahrim Youn, Kyung In Kim, Raul Rabadan, Benjamin Tycko, Yufeng Shen, Shuang Wang
BMC Medical Genomics (2018-11-06) https://doi.org/gjmhfb
DOI: 10.1186/s12920-018-0425-z · PMID: 30400878 · PMCID: PMC6218985

27. **Computational analysis reveals histotype-dependent molecular profile and actionable mutation effects across cancers**
Daniel Heim, Grégoire Montavon, Peter Hufnagl, Klaus-Robert Müller, Frederick Klauschen
Genome Medicine (2018-11-15) https://doi.org/gjmhfc
DOI: 10.1186/s13073-018-0591-9 · PMID: 30442178 · PMCID: PMC6238410

28. **CNAmet: an R package for integrating copy number, methylation and expression data**
Riku Louhimo, Sampsa Hautaniemi
Bioinformatics (2011-03-15) https://doi.org/fbq4p2
DOI: 10.1093/bioinformatics/btr019 · PMID: 21228048

29. **Impacts of somatic mutations on gene expression: an association perspective**
Peilin Jia, Zhongming Zhao
Briefings in Bioinformatics (2016-04-28) https://doi.org/gjnd5b
DOI: 10.1093/bib/bbw037 · PMID: 27127206 · PMCID: PMC5862283

30. **Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM**
Charles J. Vaske, Stephen C. Benz, J. Zachary Sanborn, Dent Earl, Christopher Szeto, Jingchun Zhu, David Haussler, Joshua M. Stuart
Bioinformatics (2010-06-15) https://doi.org/bcvgjf
DOI: 10.1093/bioinformatics/btq182 · PMID: 20529912 · PMCID: PMC2881367

31. **Systematic analysis of somatic mutations impacting gene expression in 12 tumour types**
Jiarui Ding, Melissa K. McConechy, Hugo M. Horlings, Gavin Ha, Fong Chun Chan, Tyler Funnell, Sarah C. Mullaly, Jüri Reimand, Ali Bashashati, Gary D. Bader, … Sohrab P. Shah
Nature Communications (2015-10-05) https://doi.org/f7z86p
DOI: 10.1038/ncomms9554 · PMID: 26436532 · PMCID: PMC4600750

32. **Cancer Genome Landscapes**
B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, K. W. Kinzler
Science (2013-03-28) https://doi.org/6rg
DOI: 10.1126/science.1235122 · PMID: 23539594 · PMCID: PMC3749880