Dear Editors,

We are grateful for your review of our manuscript. We found the Reviewer comments to be insightful, and the incorporation of their feedback has elevated the quality of our manuscript. The Reviewers indicated that our manuscript dealt with important subject matter and that our method has the potential to improve the analysis of pediatric cancer gene expression data. In addition to these positive remarks, the Reviewers found important areas in the manuscript that could be improved. We have addressed all of the Reviewer comments and reference changes in the resubmission.

Reviewer 1 focused on biological validation of the expression subtypes, and proposed further analysis of tumor microenvironment states in *MYCN* non-amplified (MYCN-NA) neuroblastoma tissue samples. Obtaining precious tissue samples in the pediatric oncology setting is very difficult. We went to great lengths to validate the immune/stromal infiltrate and were able to provide additional evidence to support our immune subtyping analysis. Reviewer 1 also recommended that the main text include a complete hydra analysis of a single diagnosis. We have added a complete analysis of MYCN-NA neuroblastoma and have included specific results for other diseases of interest.

In response to Reviewer 2's feedback, we have added a runtime analysis of the hydra approach compared to state-of-the-art pathway analysis tools. The hydra analysis scaled well to a large dataset. Reviewer 2 also requested a quick-start guide to help new users apply these powerful tools to their own research. We have added thorough documentation as a supplementary file to the manuscript and also uploaded the documentation to the Github repo.

We have included detailed responses to all Reviewer comments below. We are excited by the opportunity to publish our manuscript in PLoS Computational Biology and thank you for your consideration.

Sincerely,

Jacob Pfeil
UC Santa Cruz Genomics Institute
Department of Biomolecular Engineering
University of California, Santa Cruz

Response to Reviewer Comments

Typographic Conventions:

To facilitate the review of our responses, we have indexed each comment by the Reviewer and the comment number. For example, C1.1 is the first comment from the first Reviewer. The responses are similarly indexed and red in color.

Reviewer One:

C1.1: *Would the authors elaborate as to why they chose TCGA GBM data to test hydra on initially? This is a tumour predominantly found in adults and that has higher mutational frequency than pediatric brain tumours.*

R1.1: We thank the Reviewer for bringing attention to our choice of underlying expression distribution for the synthetic data generation. We chose TCGA GBM specifically because GBM is a different disease from the diseases studied in the manuscript, allowing us to avoid over-fitting the hydra method to the diseases of interest (Section: Synthetic data generation and validation). We also wanted to show the flexibility of our method by analyzing data from a variety of cancer genome sequencing projects.

C1.2: *The authors analysed 5 pediatric tumour datasets from UCSC treehouse compendium. It is very good to see that hydra and its different components can be applied to all the tumour types. However, it was confusing when performing some analyses on only 1 type of cancer or a few and not others. For continuity, would the authors consider focusing on neuroblastoma for each of the methods presented and use the other cancer types as a more supporting role? E.g. hydra filter/enrich also on neuroblastoma data.*

R1.2: We thank the Reviewer for this very helpful comment. Our method identifies many interesting patterns that warrant further investigation. We chose to report a range of analyses from different diseases to highlight the power of this approach. We agree with the Reviewer that this may be confusing to some readers and it would be better to include a complete analysis of neuroblastoma in the main text and use the other diseases in a more supporting role. We now include a complete analysis of MYCN-NA neuroblastoma with *filter*, *sweep*, and *enrich* results (Section: Hydra analysis of high-risk neuroblastoma). The *filter* analysis revealed significant multimodal expression patterns, where many of the genes were found to be druggable and may inform precision oncology applications. We also applied the *sweep* method, which identified a network of related gene sets showing multimodal expression patterns (S4 Fig). The *enrich* analysis then describes expression subtypes related to the most significantly enriched gene sets in the MYCN-NA dataset (Fig 3).

C1.3: *Have the authors considered applying hydra on the more common pediatric tumours such as lymphoma, leukemia or brain tumours? For example medulloblastoma has very well defined molecular subtypes based on transcriptomic data (methylomic, proteomic, et..)…would hydra better refine these currently well-defined subgroups by Paul Northcott, Michael Taylor, Stefan Pfister, Richard Gilbertson, etc…*

R1.3: We have hydra analyses of additional diseases ongoing, including pediatric brain tumors and leukemias, and are expecting to publish those manuscripts. We thank the Reviewer for the suggestion to apply the method to medulloblastoma and agree that it would be worthwhile for us to pursue in a future study. For this study, we chose to focus on extracranial solid tumors because they are among the most common pediatric cancers, making up 20% of all pediatric cancer diagnoses [1], and while survival rates have improved, there are few effective treatment options for the subset of patients with relapse or refractory disease [2]. Identifying expression subtypes for these diseases may improve risk stratification and discover opportunities for new therapies. These tumors also share similar histopathological features, so we hypothesized that these tumors may share similar gene expression subtypes despite significant differences in the raw expression profiles (Fig 6A, Section: Hydra analysis reveals recurrent expression subtypes across small blue round cell tumors).

C1.4: *The histological data (images and analysis) are underwhelming to support their models. For example, figure 3C would be much better supported if the authors had access to human patient samples to compare the findings of the xCell score based on hydra clusters. Immunohistochemistry for B-cells, CD8+ T-cells and fibroblasts are easy to do if samples are available. For the immune system, the spatial context is especially important when determining the 'hot' and 'cold' nature of the tumours.*

R1.4: We thank the Reviewer for this important and thoughtful comment. Unfortunately, the tissue samples described in this study were fully used for RNA sequencing, so we did not have the same tissue available to conduct the stains suggested by the Reviewer. However, the pathologist conducted additional review of the available H&E slides and was able to refine his analyses by including qualitative descriptions of the types of inflammation observed in each case whenever possible. To provide additional validation, we compared expression subtypes to state-of-the-art tumor microenvironment profiling tools. We also include survival data showing that patients with higher immune marker expression had better survival outcomes than related tumors with lower immune marker expression (Fig 7). This pattern has been observed in adult cancers and may suggest that similar immune activity is present in pediatric cancers [3].

C1.5: *The N-of-1 tumour analysis is very attractive. Rates of pediatric cancers are low which is a hurdle in facilitating better treatments for those affected. Having a more rigid histological analysis of the samples used (figure 4) is needed. General lymphocyte content is not sufficient for inflammation. The pathologist could supply additional information on neutrophil, macrophages,*

etc… to better describe the type of inflammation occurring. This level of pathological resolution is needed to better support the findings from Hydra to be utilized on n-of-1 findings.
R1.5: We thank the Reviewer for bringing this to our attention. We agree that a more detailed analysis of immune cell types and level of inflammation is needed to validate our approach. The pathologist conducted a re-review of the slides to increase the resolution of the analysis. This resulted in the identification of significant populations of mature mononuclear immune cells, plasma cells, and eosinophils. We also obtained a new sample representing cluster 3 (S5 Fig). The cluster 3 sample's H&E analysis revealed regions of necrosis, which is consistent with the wound healing signal in the expression data.

C1.6: *The hydra filter/enrich analysis of the osteosarcoma data is very interesting. I think these data would be better suited in the main figures not supplemental.*
R1.6: We thank the Reviewer for this suggestion. The osteosarcoma filter/enrich analysis of the TARGET osteosarcoma dataset revealed an important tissue signal related to infiltrating muscle cells. Pathologist review of an N-of-1 tumor sample validated this finding in an independent patient population (Fig 5).

C1.7: *For actionable drug targets, the authors highlight JAK1 signaling in Ewing Sarcoma. The authors highlight how important interpretation of the data are (Mast cell infiltration vs. the tumour) for selecting drug targets regarding the complexities of the tumour microenvironment. Can the authors supply a more concrete cancer specific actionable target to support the filter method of Hydra…something less nuanced?*
R1.7: To address this comment, we added a more quantitative analysis of the proportion of overexpressed druggable targets that may represent signal from the tumor microenvironment rather than the tumor itself. We have reviewed the druggable genes from MYCN-NA neuroblastoma. We identified 358 druggable genes by the Drug-Gene Interaction Database. Of these druggable genes, 80 correlated with an immune/stromal cell type signature (Section: Hydra analysis discovers complex tissue signatures).

C1.8: *For S4B and C, these data are impressive and should be moved to main figures.*
R1.8: We have moved these results to the main figures (Fig 4).

C1.9: *Why is S7 in the supplemental data? Did the authors assess all cluster combinations and identify only wound healing and translational regulation as important in osteosarcoma…what about cell cycle,etc..*
R1.9: We included these data in the supplemental section to limit the disease types in the main figures, but we are happy to move these results to the main figures at the Reviewer's recommendation (Fig 7). We did assess all osteosarcoma cluster combinations, but the other clusters did not have enough samples with survival outcome data ($n < 5$) to identify a correlation

with patient survival. There were enough synovial sarcoma samples with patient outcome data, so we have updated the figure to include the other survival curves (Fig 7). Additional gene expression profiles with detailed clinical data are needed to identify similar signals in the other expression subtypes.

C1.10: *Line 22: Though many accept that pediatric tumours have fewer mutations relative to adult cancers, this is not true for all pediatric cancers. Additionally, the spectrum of tumours in paediatrics and adults is remarkably different. Would the authors supply some citation for their statement or provide a comparison on mutational frequencies…even focusing on brain tumours comparing pediatrics to adults. Stefan Pfister's group in Heidelberg have published some data related to this.*
R1.10: We are grateful for the Reviewer bringing this point to our attention. We agree that a more precise description of the mutations in pediatric cancers compared to adult cancers is needed. For example, some pediatric tumors with defects in DNA repair genes respond well to checkpoint blockade therapy due to a particularly high mutation burden. We have provided a citation for our statement and have changed the language to better reflect the mutation spectrum in pediatric cancers (Line 24).

C1.11: *It may be my version, but the figures were very blurry and hard to make out details.*
R1.11: We apologize that the images were not clear in the document. We have included high-resolution figures in the main text in addition to uploading the formated images. We include another version of the manuscript without the figures in the text. There is also a link in the document to download the original high-resolution images.

C1.12: *Line 473: these findings are not too surprisingly based since the 5 cancers assessed have links back to neural crest cells.*
R1.12: We agree with the Reviewer that the findings are not overly surprising. What is notable is that existing methods for identifying relationships across samples do not identify these signals. The hydra analysis is novel in that it moves beyond the tissue-of-origin signal to reveal shared expression subtypes. We have clarified the language in the section, Hydra analysis reveals recurrent expression subtypes across small blue round cell tumors, to reflect our intended meaning.

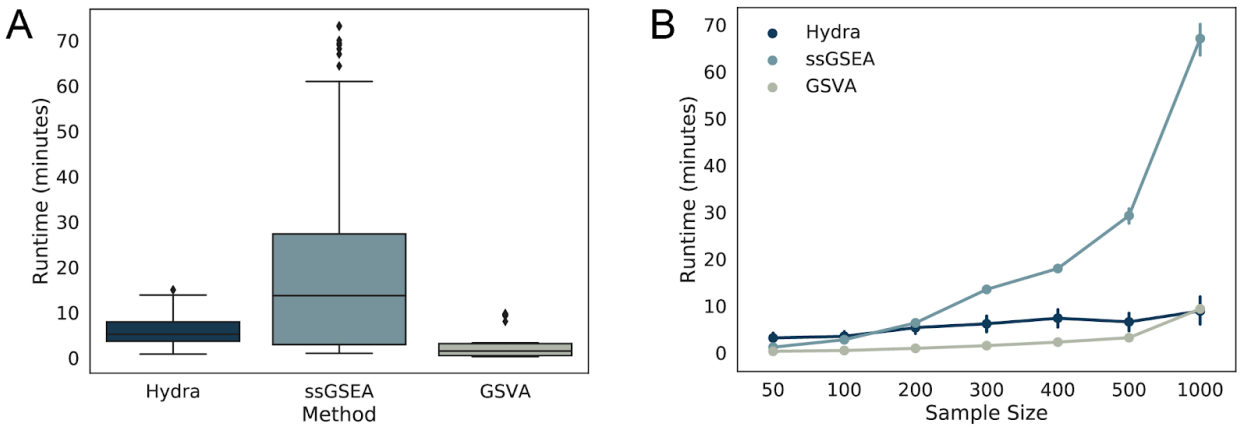
C1.13: *Line 482: should reference S7 not S6.*
R1.13: We thank the Reviewer for identifying this error. The correction has been made in the latest draft of the manuscript.

C1.14: *The method section describing the histological analysis is not sufficient. Please see above for suggestions on what to highlight.*

R1.14: We are grateful for this recommendation and have added a more detailed description of the histological analysis.

C1.15: *Several graphs in the figures have no labelled axes.*
R1.15: We apologize for the lack of sufficient axis labels. We have reviewed each graph and added detailed axes labels. In particular, S1 Fig now has appropriate axis labels.

Reviewer Two:
C2.1: *I think a discussion of runtime is warranted, especially in comparison with the other methods, and an estimation of how runtime scales with the number of samples in the analysis.*
R2.1: We thank the Reviewer for the thoughtful suggestion. We have added a discussion of runtime to the manuscript (Synthetic data generation and validation). We found that the hydra approach is computationally expensive during the training stage, but the subtype assignments for new samples are very fast once the models are trained. The GSEA-like approaches do not train the model beforehand, so the subtype assignment for new samples always takes the same amount of time. We evaluated the runtimes of ssGSEA and GSVA as compared to hydra sweep, as these two approaches are most similar conceptually to hydra *sweep*. We found that on average the hydra *sweep* approach took on average as long as ssGSEA to run, but the hydra method runtime is significantly more variable, depending on the features of the expression gene sets and the disease cohort, whereas the GSEA-based approach algorithms are independent of the biological features of the data. The GSVA approach was much faster than hydra and ssGSEA, but the predictive performance was worse than ssGSEA and hydra.



We also investigated how the runtimes scale with sample size. We repeated the above analysis, but with an effect size of 1.0, a %DEG of 25%, and a range of sample sizes, including 50, 100, 200, 300, 400, 500, 1000 samples. The hydra *sweep* and GSVA methods scaled well, but the ssGSEA runtime increased exponentially as the sample size increased (Fig 1D).

C2.2: *In the first result with synthetic data, the data is modelled "as a multivariate Gaussian distribution" (l.204), which is the running assumption in Hydra (l. 94). It would be more powerful to model the synthetic data with different assumptions and measure the performance of the method.*

R2.2: The Reviewer brings up an excellent point concerning the assumptions of the synthetic data model. While the multivariate Gaussian distribution is the standard distribution for modeling multivariate data, gene expression data can also be modeled as count data. A univariate model, like the negative binomial distribution, is an option, but the negative binomial does not account for correlations across genes, which is a necessary feature for modeling gene expression data. Multivariate count distributions, such as the multivariate Poisson, are not widely used and there is no standard approach for sampling from these distributions [4], which makes it difficult to use these distributions in practice. We agree with the Reviewer that deeper models of gene expression are needed; however, developing these methods is beyond the scope of this manuscript.

C2.3: *l.125: The URL www.dockerhub.com/jpfeil/hydra failed for me, but the correct URL seems to be https://hub.docker.com/r/jpfeil/hydra. I note that the title of the manuscript indicate that hydra is a method for "subtyping pediatric cancer cohorts", but the short description of the Docker package describes it as a "clustering pipeline to identify differentially expressed pathways". I get how those two things are related, but I believe it would be good to clarify in the text .*

R2.3: We thank the Reviewer for identifying the typo in our URL and we apologize for the inconvenience. It has been corrected in the manuscript. We also thank the Reviewer for pointing out the inconsistency between our manuscript description and the Docker package description. We have updated the package description to remedy this. Identification of differentially expressed pathways is used to identify subtypes. We have clarified this point in the text so it is clear how the tool is used to identify subtypes (Section: Hydra method).

C2.4: *Some sections are very technical. For instance, the lines 111 to 116 are referring to concepts and methods that may not be common knowledge for part of the audience. I leave it to the authors to decide if they can explain it more or prefer to leave it there, assuming that the reader will consult the references cited.*

R2.4: We thank the Reviewer for identifying this area in the manuscript that needs more explanation. We have clarified this section by removing some of the technical language, although this section remains very technical. We acknowledge that this material requires significant background knowledge. We have attempted to explain these concepts to a more general audience and we have also added additional references to other resources that the reader may consult to enhance understanding (Section: Dirichlet process gaussian mixture model).

C2.5: *I would appreciate a guide, on the Github page or even better as part of the manuscript, describing how to use the method on a completely novel cancer type and novel datasets. How to process the data, build the model and run the different tools of Hydra? Some documentation already exists, but could be expanded a bit more if the authors want more users to adopt their method.*

R2.4: We are grateful for this feedback. We have written a thorough guide for new users of the method. We have included the guide as a supplementary file (Supplementary File 2) and have published the guide on the hydra github page (https://github.com/jpfeil/hydra).

References

1. Ries L a. G, Smith MA, Gurney JG, Linet M, Tamra T, Young JL, et al. Cancer incidence and survival among children and adolescents: United States SEER Program 1975-1995. Cancer Incid Surviv Child Adolesc U S SEER Program 1975-1995. 1999 [cited 7 Jan 2020]. Available: https://www.cabdirect.org/cabdirect/abstract/20083240999
2. Ring EK, Markert JM, Gillespie GY, Friedman GK. Checkpoint Proteins in Pediatric Brain and Extracranial Solid Tumors: Opportunities for Immunotherapy. Clin Cancer Res. 2017;23: 342–350. doi:10.1158/1078-0432.CCR-16-1829
3. Fridman WH, Galon J, Dieu-Nosjean M-C, Cremer I, Fisson S, Damotte D, et al. Immune infiltration in human cancer: prognostic significance and disease control. Curr Top Microbiol Immunol. 2011;344: 1–24. doi:10.1007/82_2010_46
4. Inouye D, Yang E, Allen G, Ravikumar P. A Review of Multivariate Distributions for Count Data Derived from the Poisson Distribution. Wiley Interdiscip Rev Comput Stat. 2017;9. doi:10.1002/wics.1398