nature portfolio

Peer Review File

A multiscale functional map of somatic mutations in cancer integrating protein structure and network topology

Editorial Note: This manuscript has been previously reviewed at another journal that is not operating a transparent peer review scheme. This document only contains reviewer comments and rebuttal letters for versions considered at Nature Communications.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have thoroughly addressed all my previous comments which, in my opinion, has increased the robustness and clarity of the results and the m/s.

I am now happy to support the publication of NetFlow3D in Nat Commun.

Reviewer #2 (Remarks to the Author):

Overall the authors were highly responsive to reviewer feedback and have very significantly improved the manuscript, although several important issues remain to be carefully addressed to give confidence in the veracity and novelty of this work. I will only refer to comments not already fully addressed.

Ref2.1.2
The updated analysis very promising, but the authors need to demonstrate:
a) the observed survival correlations are independent of stage, grade, age, sex, TMB and similar features
b) these putative driver mutations are associated with downstream molecular consequences (RNA or protein abundance)
c) genes affected by these putative driver regions preferentially have phenotypes in (tumor-type-matched) CRISPR screens

Ref2.1.3
Unfortunately by showing that the search space is dramatically larger (~100x) all the authors show is that the recurrence methods focused on our hotspot methods. However that's not how practical cancer driver analyses are done. Rather, there is an assessment for each gene whether or not there are more functional mutations (nonsynonymous and stop gain/loss typically) than expected by chance. Then all functional mutations are considered putative drivers. The authors appear to have taken only a hotspot-identificiation method. It would be very helpful to be systematically reporting PPV/NPV or an equivalent pair of metrics (e.g. precision/recall).

Reviewer #2 (Remarks on code availability):

The code is very problematic in its current form. Minimal code-level documentation, a ton of hard-coded variables (e.g. line 316 of NextFlow3D.py hard-codes a list of proteins!), no unit-tests and so forth.

Reviewer #3 (Remarks to the Author):

The reviewer appreciated the extensive revision and the authors have address all my previous concerns. The revised manuscript is well-written, and the main figures are very impressive. It is deserved to be published by Nature Communications now.

Reviewer #3 (Remarks on code availability):

The reviewers can run code well.

Reviewer #4 (Remarks to the Author):

The authors have done an admirable job at responding to my previous comments from the last round of review. The new experimental validation related to PP2A is impressive. No further comments.

Point-by-Point Response Letter

– Reviewer #1 complete original comment –

Reviewer Comment	The authors have thoroughly addressed all my previous comments which, in my opinion, has increased the robustness and clarity of the results and the m/s. I am now happy to support the publication of NetFlow3D in Nat Commun.

– Reviewer #2 complete original comment –

Reviewer Comment	Reviewer #2 (Remarks to the Author): Overall the authors were highly responsive to reviewer feedback and have very significantly improved the manuscript, although several important issues remain to be carefully addressed to give confidence in the veracity and novelty of this work. I will only refer to comments not already fully addressed. Ref2.1.2 The updated analysis very promising, but the authors need to demonstrate: a) the observed survival correlations are independent of stage, grade, age, sex, TMB and similar features b) these putative driver mutations are associated with downstream molecular consequences (RNA or protein abundance) c) genes affected by these putative driver regions preferentially have phenotypes in (tumor-type-matched) CRISPR screens Ref2.1.3 Unfortunately by showing that the search space is dramatically larger (~100x) all the authors show is that the recurrence methods focused on our hotspot methods. However that's not how practical cancer driver analyses are done. Rather, there is an assessment for each gene whether or not there are more functional mutations (nonsynonymous and stop gain/loss typically) than expected by chance. Then all functional mutations are considered putative drivers. The authors appear to have taken only a hotspot-identificiation method. It would

	be very helpful to be systematically reporting PPV/NPV or an equivalent pair of metrics (e.g. precision/recall). Reviewer #2 (Remarks on code availability): The code is very problematic in its current form. Minimal code-level documentation, a ton of hard-coded variables (e.g. line 316 of NextFlow3D.py hard-codes a list of proteins!), no unit-tests and so forth.

Ref2.1.2 – Benchmarking: Functional evaluation of NetFlow3D-identified potential driver mutations **–**

Reviewer Comment	The updated analysis very promising, but the authors need to demonstrate: a) the observed survival correlations are independent of stage, grade, age, sex, TMB and similar features b) these putative driver mutations are associated with downstream molecular consequences (RNA or protein abundance) c) genes affected by these putative driver regions preferentially have phenotypes in (tumor-type-matched) CRISPR screens
Author Response	We appreciate the reviewer's positive feedback on our updated analysis, and highly value the constructive suggestions provided. Following these suggestions, we have made the following improvements: a) To exclude the effects of other clinical covariates and focus on the survival impact of NetFlow3D-identified mutations, we used a Cox regression model to evaluate the statistical association between the presence of these mutations and patient survival, with tumor stage, age, sex, and tumor mutation burden (TMB) included as covariates. This resulted in an overall drop in significance compared to the log-rank test in our previous Kaplan–Meier estimates. However, most cancer types still remained significant. Specifically, adrenocortical carcinoma (ACC) (hazard ratio [HR] = 2.71, p-value = 2.82e-2), brain low-grade glioma (LGG; HR = 1.61, p-value = 1.37e-2), thyroid carcinoma (THCA; HR = 4.87, p-value = 4.73e-2), and kidney renal clear cell carcinoma (KIRC; HR = 1.58, p-value = 4.50e-2) remained significant, while liver hepatocellular carcinoma (LIHC) lost its significance (HR = 1.21, p-value = 0.41) (**Supplementary Table 10**).

	b) We greatly appreciate the reviewer's insightful suggestion to evaluate the impact of NetFlow3D-identified mutations on downstream molecular abundance. Since our 3D clustering analysis focuses on in-frame mutations that are generally not subject to RNA-level nonsense-mediated decay, we concentrated on their impact on protein abundance. Specifically, we conducted two analyses. First, we fitted a linear regression model to evaluate the statistical association between NetFlow3D-identified mutations and protein abundance. This model controlled for gene-specific and tissue-specific baseline expression levels, as well as clinical covariates including sex, age, tumor stage, and TMB (**Supplementary Note**). This analysis revealed significant associations between the presence of NetFlow3D-identified mutations and protein abundance (p-value = 1.6e-9). As a negative benchmark, we replaced NetFlow3D-identified mutations with other mutations not identified by NetFlow3D, resulting in a non-significant association (p-value = 0.70). For a more detailed perspective, we conducted a fine-grained analysis comparing protein abundance for each gene in each cancer type, under scenarios with and without NetFlow3D-identified mutations. As a control, we repeated the analysis using other mutations. Our results revealed a significantly higher proportion of cases with differential protein abundance for NetFlow3D-identified mutations than for other mutations (**Supplementary Fig. 9**). c) We evaluated the cellular fitness effects of genes with NetFlow3D-identified mutations by leveraging core fitness genes identified from genome-scale CRISPR-Cas9 screens in 324 human cancer cell lines across 30 cancer types by Behan et al. (PMID: 30971826). In a tumor-type-matched manner, we analyzed the enrichment of these core fitness genes in NetFlow3D results. Our analysis consistently showed that genes with NetFlow3D-identified mutations were the most enriched for core fitness genes across various cancer types (**Supplementary Fig. 10**). Additionally, genes with mutations identified solely by our 3D clustering analysis showed significant enrichment in every cancer type, whereas genes without mutations in 3D clusters did not exhibit significant enrichment in any cancer type. In summary, we deeply appreciate the reviewer's insightful suggestions, which have significantly enhanced the

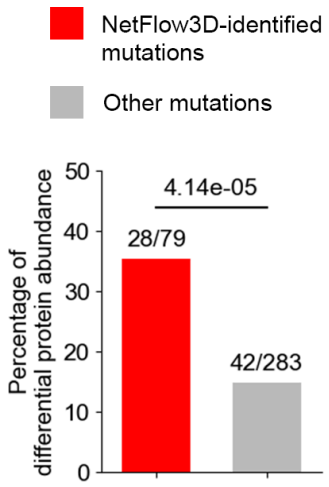
	comprehensiveness and robustness of our functional validation analyses of NetFlow3D results. We have added these analyses to our revised manuscript.
Excerpt from Revised Manuscript	[Main text; Results; Page 8] To demonstrate the clinical significance of NetFlow3D findings, we compared the overall survival between patients with somatic in-frame mutations in our preprocessed TCGA dataset, grouping them by whether their mutations were identified by NetFlow3D (Methods). We used a Cox regression model to evaluate the statistical association between NetFlow3D-identified mutations and patient survival, controlling for clinical covariates including age, sex, tumor stage, and tumor mutational burden (TMB). Our analysis revealed significant negative survival associations across multiple cancer types, including Thyroid carcinoma (THCA), Kidney renal clear cell carcinoma (KIRC), Adrenocortical carcinoma (ACC), and Brain Lower Grade Glioma (LGG) (Fig. 4d). The hazard ratios (HR) derived from the Cox model coefficients were consistently greater than 1.5 across all four cancer types (Supplementary Table 8). [Main text; Methods; Page 20] Cox regression was used to evaluate the statistical association between the presence of NetFlow3D-identified mutations and OS, with tumor stage, age, sex, and tumor mutation burden (TMB) included as covariates. For brain lower grade glioma (LGG), tumor stage was excluded from the Cox regression analysis due to unavailable data. The regression coefficients of NetFlow3D-identified mutations indicate their impact on hazard, with their exponential values representing the hazard ratio (HR) and p-values indicating the significance of the association. Furthermore, to focus on the survival impacts of mutations in the genes not curated by CGC, we treated NetFlow3D-identified mutations in known cancer genes and those not in known cancer genes as two independent predictors within a single Cox regression model. [Supplementary Table 10]

Cancer type	Hazard ratio (HR)	Coefficient	P value
Adrenocortical (ACC)	2.71	1.00	2.82e-2
Brain lower grade glioma (LGG)	1.61	0.48	1.37e-2
Thyroid (THCA)	4.87	1.58	4.73e-2
Kidney renal clear cell (KIRC)	1.58	0.46	4.50e-2

[Main text; Results; Pages 7-8] To demonstrate downstream molecular consequences of NetFlow3D-identified mutations, we evaluated their statistical association with protein abundance. Initially, we performed linear regression analysis, controlling for gene-specific and tissue-specific baseline expression levels, as well as clinical covariates including sex, age, tumor stage, and TMB (Supplementary Note). This analysis revealed significant associations between the presence of NetFlow3D-identified mutations and protein abundance (p-value: 1.6e-9). In contrast, no significant association was observed when conducting the same analysis using other mutations not identified by NetFlow3D (p-value: 0.70). For a more detailed perspective, we conducted a fine-grained analysis comparing protein abundance for each gene in each cancer type, under scenarios with and without NetFlow3D-identified mutations. As a control, we repeated the analysis using other mutations. Our results revealed a significantly higher proportion of cases with differential protein abundance for NetFlow3D-identified mutations than for other mutations (Supplementary Fig. 9). [Supplementary Note] Linear regression analysis of protein abundance TCGA proteome profiling data was obtained from Repository in the GDC data portal[1] (https://portal.gdc.cancer.gov/). Protein abundance for each gene in a sample was defined as the maximum quantification value among all its protein peptides. We used a linear regression model to evaluate the statistical association between NetFlow3D-identified mutations and protein abundance. This model controlled for gene-specific and tissue-specific baseline expression levels, as well as clinical covariates including sex, age, tumor stage, and TMB (Model: protein abundance ~ C(NetFlow3D-identified mutations) + C(gene ID) + C(cancer type) + age + C(sex) + TMB + C(tumor

stage)). As a negative benchmark, we repeated the analysis by replacing C(NetFlow3D-identified mutations) with C(other mutations).

[Supplementary Fig. 9]

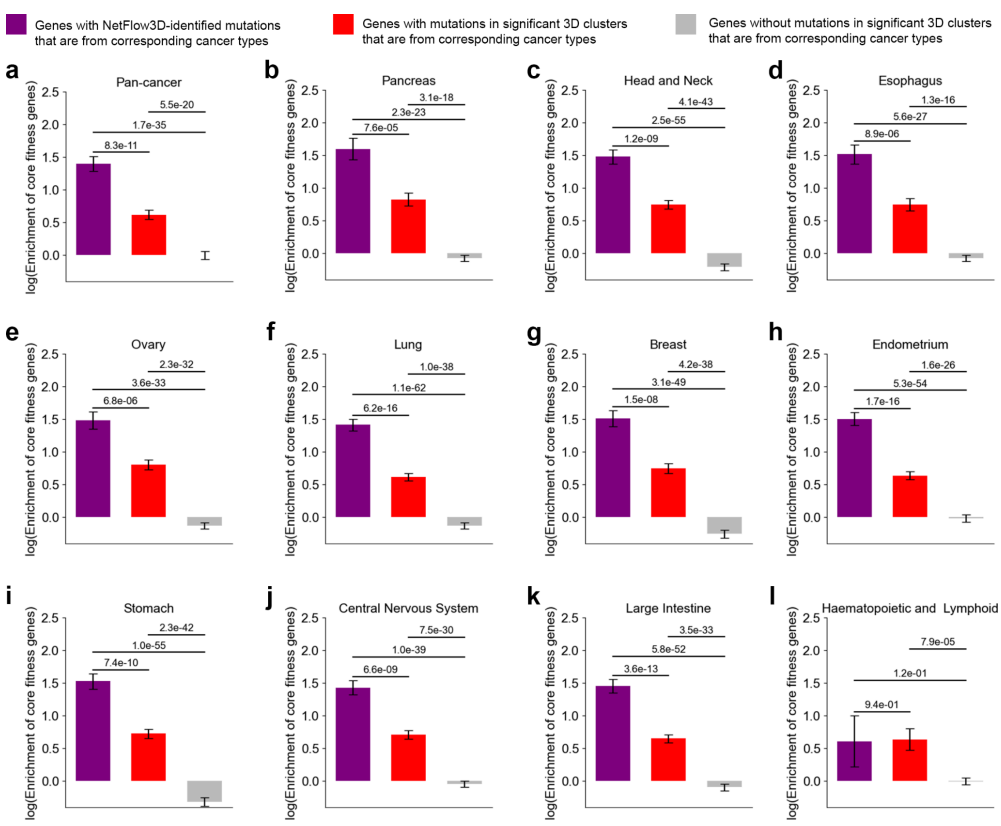


Supplementary Fig. 9. Differential protein abundance analysis per gene and cancer type in scenarios with and without mutations of interest. For each gene in each cancer type, the analysis required both scenarios to have at least 10 samples with available protein expression quantification data. The analysis was conducted for NetFlow3D-identified mutations (red) and other mutations (gray). The p-value was calculated using a two-proportion z-test.

[Main text; Results; Page 8]

To further evaluate the impact of genes with NetFlow3D-identified mutations on cellular fitness, we utilized core fitness (CF) genes identified from genome-scale CRISPR-Cas9 screens in 324 human cancer cell lines spanning 30 cancer types[60]. We analyzed the enrichment of these core fitness genes in NetFlow3D results. Our results consistently showed that, across various cancer types, genes with NetFlow3D-identified mutations are most enriched for core fitness genes (Supplementary Fig. 10). Additionally, genes with mutations identified in isolation by our 3D clustering analysis also showed significant enrichment in every cancer type, whereas genes without mutations in 3D clusters did not exhibit significant enrichment in any cancer type.

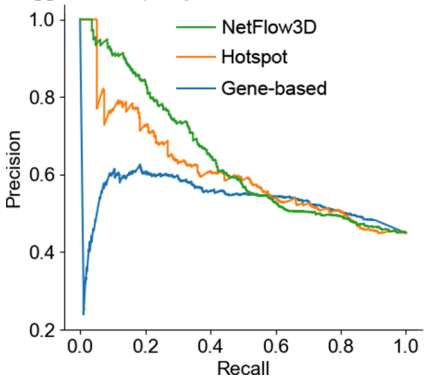
[Supplementary Fig. 10]



Supplementary Fig. 10. Analysis of core fitness gene enrichment in NetFlow3D results. Pan-cancer and cancer-type-specific core fitness genes were identified by Behan et al. using genome-scale CRISPR-Cas9 screening data in human cancer cell lines. The p-values were calculated using two-proportion z-tests.

Ref2.1.3 – Benchmarking: Benchmark NetFlow3D against recurrence-based methods –

Reviewer Comment	Unfortunately by showing that the search space is dramatically larger (~100x) all the authors show is that the recurrence methods focused on our hotspot methods. However that's not how practical cancer driver analyses are done. Rather, there is an assessment for each gene whether or not there are more functional mutations (nonsynonymous and stop gain/loss typically) than expected by chance. Then all functional mutations are considered putative drivers. The authors appear to have taken only a hotspot-identificiation method. It would be very helpful to be systematically reporting PPV/NPV or an equivalent pair of metrics (e.g. precision/recall).
Author Response	We thank the reviewer for suggesting a systematic comparison with the whole-gene-based cancer driver identification method. In response, we generated precision/recall curves for three methods

	focusing on different test units: (i) whole gene (Gene-based), (ii) 3D cluster (NetFlow3D), and (iii) single residue (Hotspot). We ensured consistency in other methodological aspects. Each method was used to rank human genes based on their statistical significance. For the hotspot and 3D clustering methods, where a gene could correspond to multiple hotspots or 3D clusters, we selected the most significant hotspot or cluster as the representative for each gene. We then generated gene-level precision-recall curves using known cancer genes from the Cancer Gene Census as positive labels and known non-cancer genes curated from the literature as negative labels. Genes with unknown labels were excluded to prevent potential misclassification. As a result, NetFlow3D demonstrated the best performance among the three methods, achieving highest precision and recall (**Supplementary Fig. 5**). We have added these validation analyses to our revised manuscript.
Excerpt from Revised Manuscript	[Main text; Results; Page 6] Beyond 3D clustering algorithms, we benchmarked NetFlow3D against other methods for identifying cancer driver mutations, including single-residue-based ("hotspot") and whole-gene-based methods. The test unit size of 3D clustering algorithms falls between these two extremes. Notably, NetFlow3D outperforms these methods, demonstrating the highest precision and recall (Supplementary Fig. 5). While the hotspot method is highly precise, it lacks power when background mutation rates are low or sample sizes are small. The whole-gene-based method, which considers the entire gene as the test unit, can dilute statistical power and lacks precision when only specific regions within the gene are responsible for driving cancer. In contrast, our 3D clustering algorithm in NetFlow3D provides flexible test unit sizes at submolecular resolution, achieving a balance of higher precision and better power. [Supplementary Fig. 5]  **Supplementary Fig. 5. Precision-recall (PR) curves of three methods for identifying cancer driver mutations.** Three methods focus on different test units: (i) 3D cluster (NetFlow3D), (ii) single residue (Hotspot), and (iii) whole gene (Gene-based). Gene-level precision-recall curves were generated using known cancer genes from the Cancer Gene Census (CGC) as positive labels and known non-cancer genes curated from the literature as negative labels. Genes with unknown labels were excluded to prevent potential misclassification. For the Hotspot and NetFlow3D methods, the most significant residue or 3D cluster was selected as the representative for each gene.

Ref2.4 – Remarks on code availability –

Reviewer Comment	The code is very problematic in its current form. Minimal code-level documentation, a ton of hard-coded variables (e.g. line 316 of NextFlow3D.py hard-codes a list of proteins!), no unit-tests and so forth.
Author Response	We thank the reviewer for the valuable insights on improving our code repository. In response, we have made enhancements in three areas: (i) comprehensive code-level documentation, (ii) increased flexibility by parameterizing inputs and configurations, and (iii) implementation and automation of unit tests. Our revised code repository can be found at https://github.com/haiyuan-yu-lab/NetFlow3D.git. First, we have improved the code-level documentation to enhance clarity and usability. This includes thorough comments and explanations within the source code, comprehensive documentation of dependencies, installation instructions, and usage guidelines, along with detailed explanations and examples for input files, arguments, and output files. Second, we have converted hard-coded variables into configurable parameters, allowing users to flexibly specify various input files and settings. For example, line 316 of NetFlow3D.py originally referenced a whitelist of proteins exempt from the RNA expression filter, following the approach by Leiserson et al. (PMID: 25501392). These proteins, encoded by 18 well-known cancer genes with low transcript detection levels, are now provided as an input file, enabling users to customize this whitelist as needed. Third, we have divided the entire NetFlow3D pipeline into seven independent modules, each of which has successfully passed unit tests. Additionally, we have automated the unit testing process in the `test_netflow3d.py` script. We truly appreciate the reviewer's constructive feedback, which has enhanced the quality, maintainability, and reliability of our code repository.

– Reviewer #3 complete original comment –

Reviewer Comment	Reviewer #3 (Remarks to the Author): The reviewer appreciated the extensive revision and the authors have address all my previous concerns. The revised manuscript is well-written, and the main figures are very impressive. It is deserved to be published by Nature Communications now. Reviewer #3 (Remarks on code availability): The reviewers can run code well.

– Reviewer #4 complete original comment –

Reviewer Comment	The authors have done an admirable job at responding to my previous comments from the last round of review. The new experimental validation related to PP2A is impressive. No further comments.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

Thank you for comprehensive & thoughtful responses, and the revised manuscript will make a significant contribution to the field.

Reviewer #2 (Remarks on code availability):

Updated code now seems highly usable.