

## ARTICLE OPEN



# Actionability classification of variants of unknown significance correlates with functional effect

Amber Johnson<sup>1</sup>, Patrick Kwok-Shing Ng<sup>1</sup>, Michael Kahle<sup>1</sup>, Julia Castillo<sup>1</sup>, Bianca Amador<sup>1</sup>, Yujia Wang<sup>2</sup>, Jia Zeng<sup>1</sup>, Vijaykumar Holla<sup>1</sup>, Thuy Vu<sup>1</sup>, Fei Su<sup>1</sup>, Sun-Hee Kim<sup>1</sup>, Tara Conway<sup>1</sup>, Xianli Jiang<sup>2</sup>, Ken Chen<sup>2</sup>, Kenna R. Mills Shaw<sup>1</sup>, Timothy A. Yap<sup>1,3</sup>, Jordi Rodon<sup>1,3</sup>, Gordon B. Mills<sup>4</sup> and Funda Meric-Bernstam<sup>1,3</sup>✉

Genomically-informed therapy requires consideration of the functional impact of genomic alterations on protein expression and/or function. However, a substantial number of variants are of unknown significance (VUS). The MD Anderson Precision Oncology Decision Support (PODS) team developed an actionability classification scheme that categorizes VUS as either “Unknown” or “Potentially” actionable based on their location within functional domains and/or proximity to known oncogenic variants. We then compared PODS VUS actionability classification with results from a functional genomics platform consisting of mutant generation and cell viability assays. 106 (24%) of 438 VUS in 20 actionable genes were classified as oncogenic in functional assays. Variants categorized by PODS as Potentially actionable ( $N = 204$ ) were more likely to be oncogenic than those categorized as Unknown ( $N = 230$ ) (37% vs 13%,  $p = 4.08e-09$ ). Our results demonstrate that rule-based actionability classification of VUS can identify patients more likely to have actionable variants for consideration with genomically-matched therapy.

npj Precision Oncology (2023)7:67; <https://doi.org/10.1038/s41698-023-00420-w>

## INTRODUCTION

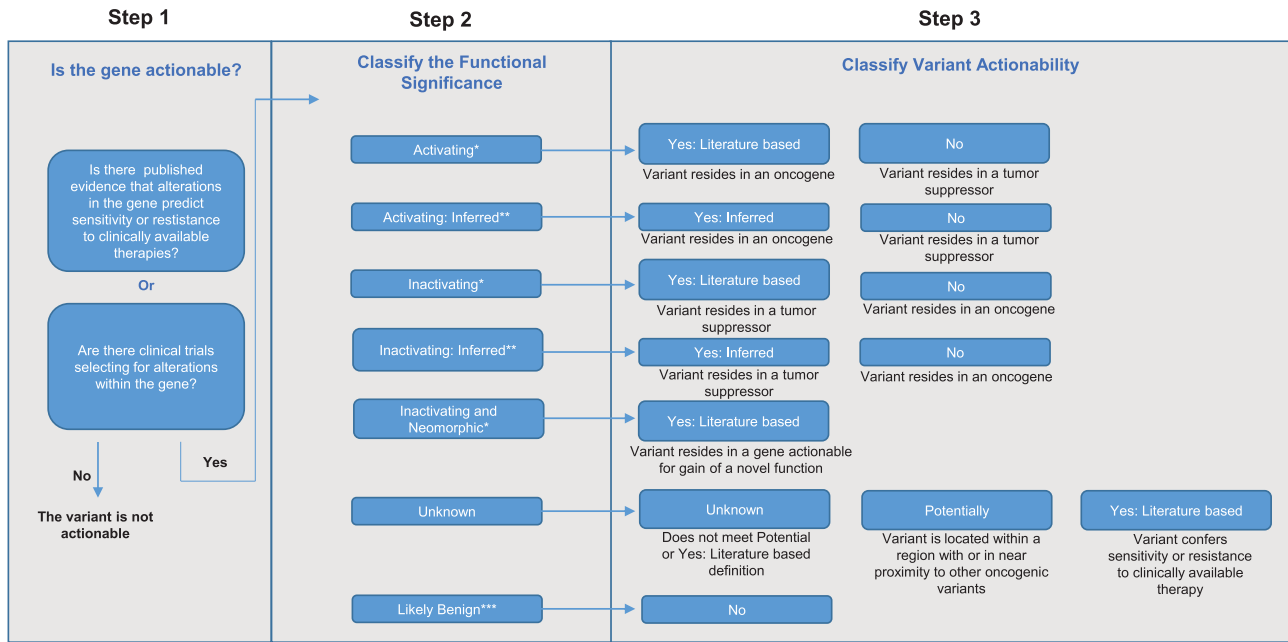
Genomic sequencing is often performed in patients with advanced or metastatic disease in order to identify alterations that may affect therapeutic decision-making and provide additional approved or investigational options. However, not all patients have alterations in actionable genes, and furthermore, not all alterations in actionable genes affect gene function. Current standards and guidelines for delivering tumor genomic sequencing reports within a clinical setting include the requirement for interpretation and categorization of detected variants for their clinical significance<sup>1</sup>. A joint consensus recommendation by the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists recommended a four-tiered system designating variants of strong (tier 1), potential (tier 2), unknown (tier 3), and benign (tier 4) clinical significance<sup>2</sup>. Likewise, the FDA recently published a fact sheet detailing three levels of evidence for tumor biomarkers detected within next-generation sequencing tests: companion diagnostic (level 1), clinical significance (level 2), and potential clinical significance (level 3)<sup>3</sup>, and the European Society for Medical Oncology (ESMO) has published the ESMO Scale of Clinical Actionability for molecular Targets (ESCAT)<sup>4</sup>, along with data showing a correlation between improved response and the ESCAT tier<sup>4,5</sup>. Several academic groups have also published their own schemes for determining the level of evidence for targeting a specific genomic alteration with a particular therapy within an indicated tumor type<sup>4,6–20</sup>, and multiple knowledgebases exist to serve as a source of alteration-level interpretation data (e.g., PersonalizedCancerTherapy.org, OncoKB, The Jackson Laboratory Clinical Knowledgebase, as previously reviewed<sup>21</sup>). While these knowledgebases provide essential information for identifying alterations of significance, a large percentage of alterations detected in patient samples have not been previously

experimentally or clinically characterized, may not appear within these knowledgebases, and fall within the unknown or uncertain classification (variants of unknown significance, VUS).

In a previous study, we quantified the number of VUS within therapeutically actionable genes identified in patients' genomic sequencing reports reviewed by the PODS team<sup>22</sup>. 48% of variant annotations provided to oncologists indicated that the variant is a VUS. The large percentage of VUS identified within patients' sequencing reports presents a great challenge to clinicians. Thus, many groups have developed high-throughput pipelines to characterize functionally somatic and/or germline VUS<sup>23–30</sup>, including a functional genomics platform established at MD Anderson<sup>31</sup>. This platform utilizes two cell lines, MCF10A and Ba/F3, to measure an alteration's impact on cell viability under growth factor independent conditions. While these platforms generate invaluable information with regards to the clinical actionability of a specific variant, a bottleneck still exists in testing and generating these data in a timely enough manner for point-of-care decision making.

To address the real-time need for determining whether a variant is likely to be functionally significant and therapeutically actionable, the PODS team created a tiered actionability scheme<sup>32</sup> (Fig. 1). The first step in the scheme is to determine if the gene harboring the variant is therapeutically actionable. PODS scientists classify genes as therapeutically actionable if there is at minimum preclinical evidence that alterations within the gene predict sensitivity or resistance to a clinically available therapy (FDA-approved or investigational agent available within clinical trials) or if alterations within the gene are part of current clinical trial eligibility criteria. Variants within actionable genes are then researched for any known or predicted (inferred) functional impact or therapeutic relevance. Based on these data, PODS assigns a Functional Significance value, which is then utilized to

<sup>1</sup>Institute for Personalized Cancer Therapy, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>2</sup>Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>3</sup>Investigational Cancer Therapeutics (Phase I Program), The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>4</sup>Division of Oncological Sciences, Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA. ✉email: [fmeric@mdanderson.org](mailto:fmeric@mdanderson.org)



**Fig. 1 Variant Actionability classification.** Schematic depicting PODS' process of defining variant actionability.

determine Variant Actionability, which may be captured as Yes (based on published literature, inferred due to the loss of characterized domains, or based on functional genomics testing), Potentially, Unknown, or No. For example, activating mutations in the oncogene *BRAF* and inactivating mutations in the tumor suppressor gene *BRCA1* are considered actionable (Supplementary Table 1). If the variant's Functional Significance is Unknown and there is no known effect of the variant on therapeutic sensitivity or resistance, the variant is classified as either Unknown or Potentially for its Variant Actionability. Variants are Potentially actionable if they are located within a functional domain where other oncogenic variants are known to occur, or, if in general, are within close proximity to other oncogenic variants. Otherwise, these variants are categorized as Unknown for actionability.

In this study, we sought to determine how often VUS in actionable genes are oncogenic using a functional genomics platform that utilizes two cell lines, MCF10A and BaF3, to measure an alteration's impact on cell viability under growth factor independent conditions<sup>31</sup>. We then determined whether our actionability framework that further classifies VUS into Potentially actionable or Unknown actionability enriches the knowledgebase with variants more likely to be functionally significant, and thus, of therapeutic relevance.

## RESULTS

### Functional characterization of VUS identified by PODS

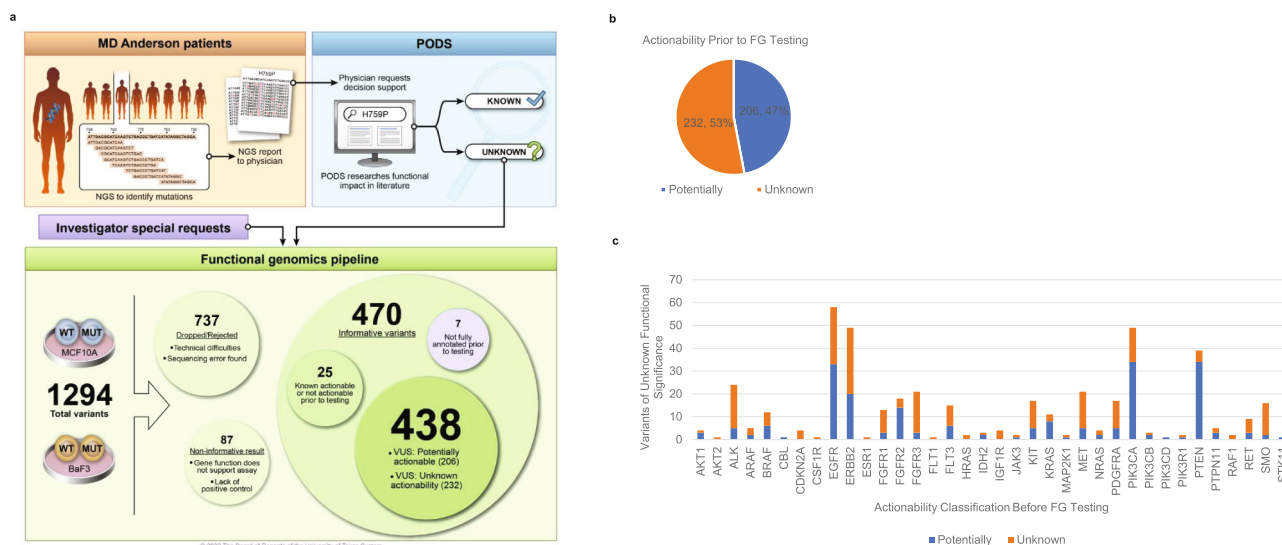
Informative results were obtained for 470 variants requested for functional genomics testing. 438 of these variants (Fig. 2a) were annotated prior to functional genomics testing as Unknown for Functional Significance and either Potentially ( $N=206$ , 47%) or Unknown ( $N=232$ , 53%) for Variant Actionability (Fig. 2b). The remaining 32 variants were either not annotated prior to testing ( $N=7$ ) or they were already known to be actionable or not actionable based on data curated from the published literature ( $N=25$ ). For all 438 variants of unknown functional significance, variants also Unknown for actionability spanned 33 of 36 genes

tested, whereas those Potentially actionable spanned 28 of 36 genes tested (Fig. 2c).

Of the 438 VUS, 106 (24%) increased cell viability in at least one cell line in comparison with its wildtype counterpart (oncogenic), 328 (75%) had no effect differing from wildtype in either cell line or decreased cell viability in comparison with the wildtype (not oncogenic), and 4 (1%) had opposing effects within the two cell lines (conflicting data) (Fig. 3a). After VUS were submitted to the functional genomics platform, new literature was found before functional genomics testing was completed for 12 variants, including 10 (2.3%) variants known to be actionable and 2 (0.5%) known to be not actionable due to newly curated literature (Fig. 3b). Of the 10 known to be actionable variants (Supplementary Table 2), 7 demonstrated a gain-of-function within the published literature that was also observed within the platform. The remaining three variants had no published functional data but were nonetheless considered actionable by PODS due to drug sensitivity or resistance data. The two variants known to be not actionable showed no effect within the functional genomics platform. Eight variants could not be clearly classified as actionable or not actionable due to either cell type-dependent functional effects (4 variants, Fig. 3a) or conflicting data between functional genomics results and the published literature (4 variants, Supplementary Table 3). Thus, 97 (22.1%) variants became actionable and 321 (73.3%) became not actionable due solely to functional genomics data (Fig. 3b).

### PODS classification of VUS as potentially actionable correlates with functional characterization

Next, we determined if those variants categorized as Potentially actionable by PODS classification were more likely to be confirmed actionable by functional genomics testing. The four variants that had opposing effects in the two cell lines tested were not included further. 30/230 (13%) variants categorized as Unknown for actionability prior to testing were found to be functionally oncogenic in at least one of the two cell lines tested; whereas, 76/204 (37%) variants categorized as Potentially



**Fig. 2 Functional genomics screening of VUS.** 1294 variants requested for functional genomics (FG) testing during the years 2015–2019 by the PODS team entered the pipeline for construction of mutant-expressing clones. 737 variants were either dropped from the pipeline due to technical difficulties or subsequently rejected due to a sequencing error found upon validation. Informative results for ingestion within the PODS knowledgebase were obtained for 470 variants. Of these, 438 were fully annotated prior to the functional genomics testing and determined to be variants of unknown functional significance classified as either Unknown ( $n = 232$ ) or Potentially actionable ( $n = 206$ ). Results were excluded for 87 variants deemed non-informative due to the lack of an informative positive control or the function of the gene did not support actionability assessment within the assay. Copyright used with the permission of The Board of Regents of the University of Texas System through The University of Texas MD Anderson Cancer Center (**a**). Percentage of 438 VUS pre-classified as either Unknown or Potentially actionable (**b**) and represented per gene (**c**).

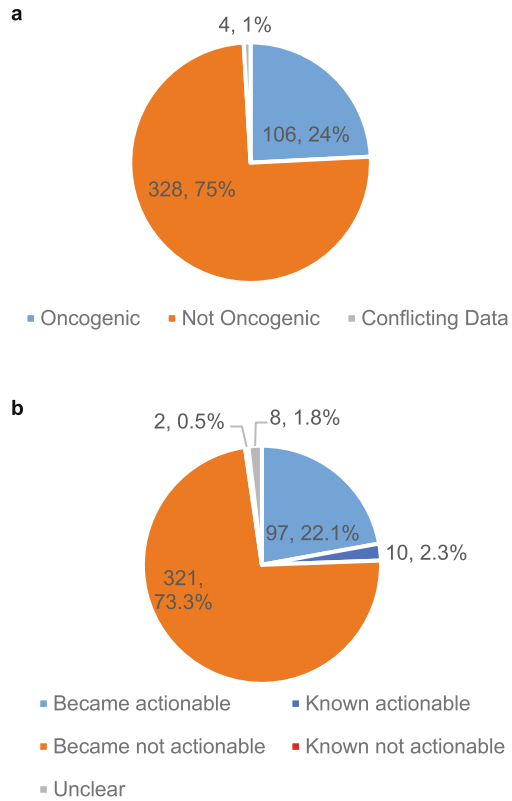
actionable prior to testing were found to be functionally oncogenic within the functional genomics platform (Fig. 4a). Thus, those annotated as Potentially actionable are more likely to be functionally validated as actionable (Fisher's Exact Test, odds ratio: 3.94,  $p = 4.08e-09$ ) than those classified as Unknown for actionability. We also computationally applied the PODS Unknown/Potentially actionability classification scheme to a second collection of variants with informative functional genomics results whose testing originates from submission to the platform by other groups (not PODS) or from our team but without prior annotation. Of 777 variants, 6590 were categorized as Potentially actionable based solely upon the described PODS criteria, while 118 were categorized as Unknown. 44% ( $n = 290$ ) of the Potentially actionable variants were verified to be oncogenic in the functional genomics platform compared with only 8% ( $n = 9$ ) of those categorized as Unknown (Fisher's Exact Test, odds ratio: 9.50,  $p = 4.719e-16$ ; Fig. 4b). Thus, these data support our conclusion that variants annotated by PODS as Potentially actionable are more likely to be oncogenic than those annotated as Unknown for actionability.

VUS that were oncogenic within the functional genomics platform ( $N = 106$ , Supplementary Table 4) pre-categorized as Unknown for actionability spanned 13 genes, and those pre-categorized as Potentially actionable spanned 16 genes (Fig. 5a). We next determined proximity for the nearest actionable alteration for all 106 variants determined to be functionally oncogenic in the functional genomics platform and pre-categorized as either Unknown or Potentially actionable. Two variants were excluded as they are truncating mutations, which are typically assigned an actionability value based on what is known in the published literature regarding the functional impact of the lost protein region and not solely on its proximity to other known actionable alterations. Of those 104 variants examined, there is an actionable alteration at the same amino acid position or within the amino acid span (for in-frame insertions and/or deletions) for 52% of the variants (Fig. 5b). For another 23% of

variants, another actionable alteration exists within at least 2 amino acids. Thus, the majority of VUS demonstrated to be oncogenic within the functional genomics platform are within 2 amino acids of another alteration also demonstrated to be actionable based on the published literature or functional genomics testing.

## DISCUSSION

Somatic genomic sequencing is recommended for all patients when one or more genomic biomarkers are linked to a regulatory body-approved therapy in the patient's tumor type<sup>33</sup>. These genomic alterations are designated as AMP Tier 1A/ PODS level 1A<sup>34</sup> and have the highest level of evidence for clinical action. If the drug approval is linked to a specific alteration, such as *BRAF* V600E, interpreting the results and therapeutic choices is relatively straight forward. However, some FDA indications and professional guidelines are linked to a general type of alteration and not a specific variant. For example, erdafitinib is FDA approved for the treatment of urothelial cancers with susceptible *FGFR3* or *FGFR2* genetic alterations, per FDA label<sup>35</sup>. In this case, functionally characterizing novel *FGFR2* and *FGFR3* alterations has significant therapeutic implications. For patients where no Tier 1A alterations are detected or they were previously acted upon, identifying functionally significant alterations that may be predictive of response to targeted therapies investigated within the context of a clinical trial becomes equally as important. However, a large portion of cancer-associated mutations have not been functionally characterized. Among all 16,738 annotated alterations within PODS as of 4/27/2022, 65% are not known for their therapeutic actionability (Unknown or Potentially actionable, Supplementary Fig. 1). Moreover, in a previous study, we determined that approximately 50% of 535 patients assessed by the PODS team had no clearly actionable mutation to pursue for enrollment on a clinical trial at the time of assessment<sup>22</sup>, and a similar study found



**Fig. 3 Actionable variants identified from functional genomics screening.** Variants of unknown significance were categorized as oncogenic or not oncogenic dependent upon whether the mutation increased cell viability in comparison with expression of the wildtype gene in at least one tested cell line (oncogenic) or either decreased cell viability or conferred no change in cell viability in comparison with the wildtype (not oncogenic). Variants that displayed opposing effects (increased and decreased) in the two cell lines were categorized as conflicting data (a). Knowledge accumulated from the published literature within the PODS knowledgebase at the time of functional genomics results were returned was assessed. Variants where literature-based actionability already existed are “Known actionable”; variants where literature-based data existed reflecting a non-actionable assertion are “Known not actionable”; and those that became actionable or not actionable due solely to functional genomics data are shown. If conflicting data existed, the variants are classified as “Unclear” (b).

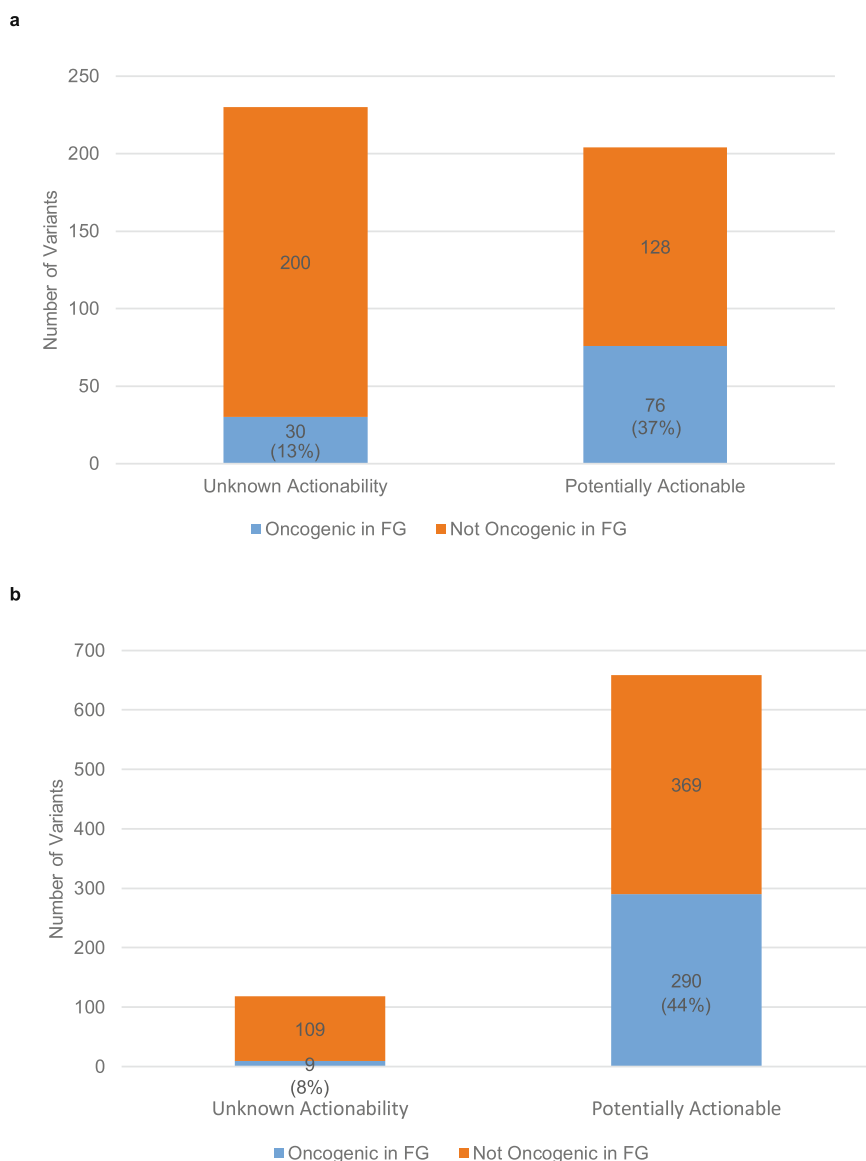
only 41% of patient samples had a potentially actionable mutation<sup>10</sup>.

The ideal scenario for determining whether a VUS is likely to be actionable is to experimentally test the function of the alteration. Functional genomics platforms are one way to characterize the tumorigenic potential of a large number of mutations. The platform utilized in this study measures a mutation’s impact on cell viability in growth factor independent conditions compared with expression of its wildtype counterpart. 24% of mutations tested increased cell viability (Fig. 3a), providing evidence that they may be tumor-promoting events that could potentially confer sensitivity to targeted therapies. For the remainder of the variants, these may either be benign passenger mutations, or their tumorigenic properties may not be seen in the genetic background of the cell models used and/or the assay setting of the platform. For example, we observed several mutations within *FGFR2* that confer increased survival in the presence of FGF ligand, but not in its absence (data not shown). Additionally, we acknowledge another limitation of the current platform. Some

variants may promote other tumorigenic phenotypes such as migration or angiogenesis, which are not assessed on this platform. Therefore, we excluded variants (i.e., annotated as non-informative) residing in genes where neither the wildtype nor any variation of the gene promoted cell viability to avoid over interpretation of the testing result.

While the functional analysis data are value-adding for variant-level knowledgebases, such as PODS, the information is typically not generated quick enough to influence care for the initial patient for which the alteration was identified. Notably, the PODS functional genomics effort was initiated with the intent to be able to guide decision-making for individual patients. However, as more of the advanced cancer population underwent comprehensive testing on platforms that go beyond “hot spot” testing for recurrent mutations, it quickly became apparent that the current genomics platform does not have fast enough turn-around to guide the care of individual cancer patients who often have rapidly progressing disease. Therefore, instead, we embarked on systematic characterization of recurrent VUS in known drivers, in order to impact subsequent patients with these mutations. In the future, tracking functional impact of individual mutations shared from larger scale functional genomics efforts, as well as tracking individual clinical outcome data of patients with genomic alterations treated on genomically-informed trials, will likely improve decision support efforts.

Methods for predicting the likelihood that an alteration is tumor-promoting are value-adding when functional data is not available. Multiple informatics tools, such as Mutation Assessor<sup>36</sup>, Hotspot3D<sup>37</sup>, HotMAPS<sup>38</sup>, SIFT<sup>39</sup>, Polyphen-2<sup>40</sup>, FATHMM-XF<sup>41</sup>, CanPredict<sup>42</sup>, MutationTaster<sup>43</sup>, SNAP<sup>44</sup>, GAVIN<sup>45</sup>, EVE<sup>46</sup>, CGI<sup>47,48</sup>, VEST4<sup>49</sup>, CScape<sup>50</sup>, and CHASM<sup>51</sup> were developed for this purpose. These tools use a variety of properties and features to predict the functional impact of a mutation, including evolutionary conservation, protein features, 3D protein structures, machine learning from curated driver mutations, and other codon-specific physiochemical properties. Some tools such as CanDrA<sup>52</sup> combine features across tools to make a prediction. Another tool, e-MutPath, assesses the effect mutations have on functional pathways by overlapping gene expression perturbations in cancer with patient-specific mutations and identified perturbations in protein-protein interactions<sup>53</sup>. With so many tools and options that may give varying predictions, it can be difficult to discern the best approach, although various comparisons have been made<sup>31,52,54,55</sup>. We chose four widely used prediction tools in order to assess how they perform relative to our functional genomics results (Supplementary Fig. 2). Alterations predicted to be drivers by CGI<sup>47</sup> (30% vs 10%; Fisher’s Exact Test, odds ratio: 4.08,  $p = 3.221e-06$ ), VEST4<sup>49</sup> (35% vs 12%; Fisher’s Exact Test, odds ratio: 3.99,  $p = 4.187e-08$ ), CHASMplus<sup>56</sup> (30% vs 6%; Fisher’s Exact Test, odds ratio: 6.23,  $p = 9.849e-09$ ), and CScape<sup>50</sup> (24% vs 4%; Fisher’s Exact Test, odds ratio: 7.71,  $p = 0.01449$ ) were more likely to be oncogenic in the functional genomics platform than those predicted by the respective tools to be passengers. However, with large-scale decision support efforts, the only known input may be the amino acid change, limiting the application of some bioinformatic prediction tools. Indeed, not all alterations could be called by each tool. With only protein amino acid change as input, 360/434 (83%) were called by CGI, 401/434 (92%) were called by VEST4, 401/434 (92%) were called by CHASMplus, and 373/434 (86%) were called by CScape. Thus, the PODS team’s schema is complimentary to these tools. It relies on curated knowledge of known functional mutations likely to be oncogenic in conjunction with manual assessment of protein domains to classify a VUS as Potentially actionable if it resides within the same functional domain as other oncogenic mutations and/or is located in close proximity or at the same codon as other oncogenic mutations. This approach is supported by literature



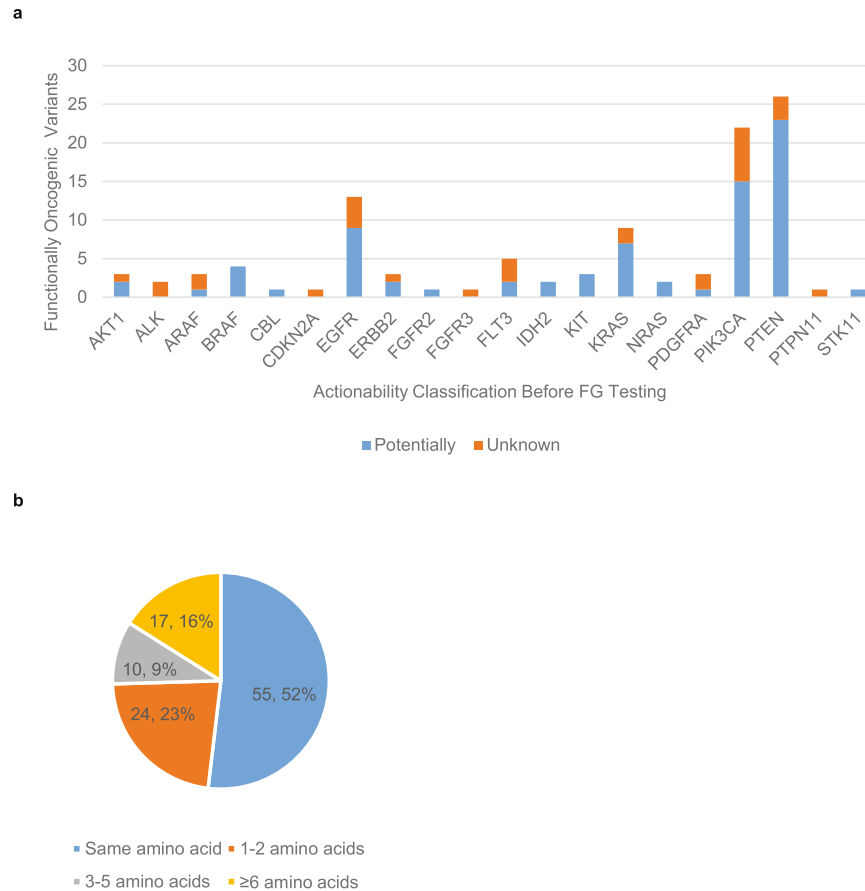
**Fig. 4 Potentially actionable variants are more likely to be functionally validated.** Variants pre-classified as either Unknown or Potentially actionable prior to functional genomics (FG) testing are displayed by whether they were oncogenic (increased cell viability) or not oncogenic (no change or decreased cell viability) within the FG pipeline (Fisher's Exact Test, odds ratio: 3.94,  $p = 4.08e-09$ ) (a). A second set of variants computationally determined to be Potentially actionable or Unknown for actionability are displayed by whether they were oncogenic (increased cell viability) or not oncogenic (no change or decreased cell viability) within the FG pipeline (Fisher's Exact Test, odds ratio: 9.50,  $p = 4.719e-16$ ) (b).

demonstrating that functionally significant, non-frameshift/truncating alterations tend to cluster in specific functional regions of the gene. For example, 17/20 of the most frequent *PIK3CA* mutations in breast cancer<sup>57</sup> that are also oncogenic, reside within a region characterized and captured within UniProt<sup>58</sup>.

Additionally, if an uncharacterized mutation is located at a hotspot, defined as a recurrently mutated amino acid in cancer, it may be considered more likely to be pathogenic. Hot spot annotation databases can be useful for predicting functional effect<sup>59</sup>. For example, the hotspot *KRAS* codon G12 is substituted for a variety of other amino acids within cancer samples. Many of these substitutions have been shown to be oncogenic and/or confer the same functional effect of impairing hydrolysis of GTP, (A<sup>60,61</sup>/C<sup>61,62</sup>/D<sup>61</sup>/F<sup>63</sup>/R<sup>61</sup>/S<sup>60,64</sup>/V<sup>61</sup>/Y<sup>65</sup>), albeit to differing degrees. Thus, other non-characterized variants of G12 would be

considered Potentially actionable. However, other recurrently mutated codons are polymorphisms and benign in nature, such as *KIT* M541L (rs3822214, dbSNP). Thus, the PODS team does not rely on frequency of detection to differentiate between Potentially actionable and Unknown for actionability variants. Our approach necessitates that other alterations at the hotspot or functionally characterized region alter protein function in a manner that is likely tumor promoting in order for VUS at that codon or region to be classified as Potentially actionable.

Until this study, the merit of our tiered actionability scheme for VUS had not been tested for the value of a Potential call. Our data here show that alterations categorized as Potentially actionable by the criteria described are more likely to be functionally significant than those categorized as Unknown for actionability (Fig. 4), as demonstrated in cell viability assays. We also demonstrate that the majority of the functionally validated variants are in near proximity



**Fig. 5 Functionally validated oncogenic variants are in near proximity to other oncogenic variants.** Variants pre-classified as either Unknown or Potentially actionable prior to functional genomics (FG) testing and demonstrated to be oncogenic within the FG platform are shown per gene (**a**). The proximity to other known actionable alterations, as determined by accessing the PODS knowledgebase on 9/17/2021, is shown (**b**).

(1–2 amino acids) to other oncogenic variants (Fig. 5b). These data suggest that among the Potentially actionable variants, there may be even more stratification of likelihood that is useful: those within 2 amino acids of another oncogenic variant being even more likely to be functionally oncogenic.

There has been some debate about how to optimize efficacy signal in genomically-informed trials. In our study, of the 438 VUS, only 24% were oncogenic. This supports the idea that when genomically-informed trials are conducted, if the goal is to enhance the efficacy signal, accrual either should be limited to known alterations, or alternately incorporate a functional annotation step that will incorporate emerging alterations with literature support and provide a tiered classification of VUS for consideration of enrollment in selected scenarios.

Altogether these data demonstrate that functional annotations relying on experimental data cannot be replaced by predicted functionality by proximity and protein features, as 63% of VUS classified as Potentially actionable were not functionally validated in the systems assessed (Fig. 4). However, the PODS tiered VUS actionability scheme does add value in stratifying alterations more likely to be functionally significant: 37% of the Potentially actionable variants had a functionally significant effect in the functional genomics platform. This information would be important to take into consideration for an individual patient along with expected therapeutic efficacy of the genomically-matched therapy and other treatment options available. Therefore, genomic annotation of VUS may identify additional patients that benefit from emerging therapeutics.

## METHODS

### Clinical genomic testing and PODS variant annotations

Patients underwent genomic testing using local or commercial clinical genomic next-generation sequencing tests as standard of care or under genomic sequencing studies with written informed consent (NCT01772771). The prospective genomic testing protocol (with written informed consent), as well as a protocol for retrospective review of clinical genomic testing results (with waiver of informed consent) was reviewed and approved by the MD Anderson Cancer Center Institutional Review Board. Variants identified within patients' CLIA sequencing reports are entered into the PODS knowledgebase. PODS scientists classified genes as therapeutically actionable (Fig. 1)<sup>32</sup>. Variants within actionable genes were then researched for any known or predicted functional impact or therapeutic relevance. Based on these data, PODS assigned a Functional Significance classification, which was then utilized to determine the Variant Actionability classification. Some variants may have more than one Variant Actionability value; each value associated with treatment or resistance to a specific drug or class of drugs (Supplementary Table 1). In these cases, the highest value was utilized for all analyses within the paper (Yes > Potentially > Unknown > No).

We also computationally applied our rules for assignment of an Unknown or Potentially actionable value to a second set of variants, which originated from other groups also utilizing the functional genomics platform or variants submitted by PODS without prior annotation (Fig. 4b). Like with manual annotation, a

Potentially actionable value is given if either the alteration resides within a protein feature considered functional (disordered regions excluded) and that contains at least one actionable mutation of the subtype missense, in-frame insertion, in-frame deletion, duplication, or deletion-insertion within or overlapping with the amino acid range of the alteration, or the alteration is within five amino acids of an actionable alteration of the subtypes previously specified irrespective of location within a functionally characterized protein feature.

### Variant submission and testing

1,294 variants requested for testing by the PODS team during the years 2015–2019 entered the functional genomics pipeline (Fig. 2a). Lentiviral vectors, originating from Clontech, expressing variants of interest or corresponding wildtype were constructed and validated as previously described<sup>29</sup>. Seven hundred thirty-seven variants were dropped out during the process due to various technical reasons, including unavailability of correct ORF and failure in full-length sequencing validation. Expression vectors for 557 variants were constructed, full-length sequencing validated, and functional testing proceeded with two growth factor-dependent cell line models, Ba/F3 and MCF10A, as previously described<sup>31</sup>. Briefly, lentivirus vectors expressing either the wildtype gene or the variant of interest were expressed within the two cell lines. Ba/F3 cells originate from MD Anderson Characterized Cell Line Core facility, and MCF10A cells originate from ATCC (CRL-10317). Transduced cells were incubated without dependent growth factors (i.e. IL-3 for Ba/F3, EGF and insulin for MCF10A) for 3 weeks. Cell viability was measured during the 3-week assay period, and the effect of the variant was compared with the corresponding wild type. Results were considered informative (470 variants) for utilization within the PODS knowledgebase if expression of at least one variant of the gene or the wildtype gene promoted cell viability within the cell line; thus, demonstrating that the oncogenic potential of the gene can be observed in the genetic background of the cell model utilized. Otherwise, results were considered non-informative in the respective cell line(s). 87 variant results were deemed non-informative for this reason or because the wildtype gene functioned in a manner opposite of the effect being examined for actionability. Specifically, expression of *FGF6*, typically considered an oncogene<sup>66,67</sup>, suppressed cell growth in the assay; and *PTCH1*, typically considered a tumor suppressor gene<sup>68,69</sup>, increased cell viability. Thus, we could not confidently assess mutations for actionable gain-of-function (*FGF6*) or loss-of-function (*PTCH1*) mutations. *ARAF* mutations were only considered non-informative within MCF10A cells, as the wildtype gene demonstrated tumor suppressive activity within this cell line but not Ba/F3 cells. *ARAF* is typically considered an oncogene<sup>70,71</sup>; thus, oncogenic gain-of-function mutations within MCF10A cells could not be determined.

### Determining functionally validated, actionable variants

Variants were considered actionable by functional genomics testing if they increased cell survival and/or proliferation in comparison with the wildtype gene in at least one cell line tested, and the variant resides within a gene classified as actionable by the PODS team at the time the result was captured. At the time of data capture, four variants remained Unknown for Functional Significance after functional genomics testing after considering their effect within the platform in combination with what was known at the time within the published literature. Detailed annotations are provided in Supplementary Table 3 for these variants. A two-sided Fisher's Exact Test was performed to determine if those annotated as potentially actionable were more likely to be functionally validated as actionable.

### Utilizing bioinformatic prediction tools

The publicly available web interface for the Cancer Genome Interpreter<sup>47</sup>, available at <https://www.cancergenomeinterpreter.org>, was utilized to categorize variants as a driver (predicted and/or annotated) or a passenger. Additionally, OpenCravat<sup>72</sup> was utilized to access prediction tools VEST4<sup>73</sup>, CHASMplus<sup>56</sup>, and CScape<sup>50</sup>. First, protein amino acid changes were mapped to a genomic position and DNA change via TransVar<sup>74</sup> version 2.5.10.20211024 using UCSC reference genome HG19 that was indexed by samtools (version 1.17)<sup>75</sup>. Outputs from TransVar were checked against input protein changes, and incorrect mappings were removed. For variants with multiple inferences, results were retained for those where the cDNA and protein amino acid change match between TransVar and OpenCravat. OpenCravat inferred variants with the most severe sequence ontology as primary variants and retained other mapped transcripts as secondary information. Alterations, where the inferred amino acid change from the prediction tool does not match with the amino acid change serving as input from functional genomics testing, were discarded from inclusion. For inferences with primary transcripts or protein changes that did not match between TransVar and OpenCravat, we selected the MANE and Ensembl transcripts with the same protein changes for CHASMplus and VEST4. As no transcript information for CScape was available in OpenCravat, we excluded predictions for those unmatched from the analysis. A significance level of 0.05 was used for determining predicted oncogenic drivers versus passengers for CHASMplus and VEST4, and a threshold of 0.5 was used for CScape as recommended by the tool.

### Determining the nearest oncogenic variant

To determine the nearest oncogenic variant to a “variant of interest” that was validated to be actionable by functional genomics testing, a search of the PODS database was conducted on 9/17/2021. Variants that had at least one actionability value of “Yes” based on published literature or functional genomics testing and are of the subtype missense, in-frame deletion, in-frame insertion, duplication, or deletion-insertion qualified as “other” oncogenic variants. The distance between the two variants was calculated as follows:

- When the variant of interest represents a single codon and the other oncogenic variant represents a single codon, the difference between the two codons was subtracted (e.g. D323A and D323E; distance = 0)
- When the variant of interest represents a single codon and the other oncogenic variant comprises multiple codons, the distance was determined to be 0 if the variant of interest (e.g. K385M) resides within the amino acid range of the other oncogenic variant (e.g. Y375\_K455del)
- When the variant of interest comprises multiple codons and the other oncogenic variant represents a single codon, the distance was calculated to be 0 if the other oncogenic variant (e.g., Y65C) resides within the range of amino acids for the variant of interest (e.g., H64\_Y65\_delinsQS).
- When the variant of interest comprises multiple codons and the other oncogenic variant also comprises multiple codons, the distance was calculated as 0 if the amino acid range of the two variants are identical or the amino acid range of either the variant of interest or the nearest oncogenic variant is nested within the other's amino acid range (e.g., P551\_M552 > L and K550\_K558del). For all other scenarios, the distance is calculated as the difference between the two most N-terminal amino acids (e.g., D770\_N771insGF and N771\_P772insH, distance = 1)

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

Patients' tumors were sequenced within a very large variety of external CLIA-certified laboratories, in addition to MD Anderson's internal CLIA-certified laboratory. Clinical sequencing data was collected and entered within an internal MD Anderson database as part of the informed consent protocol (NCT01772771) from these various sources. The PODS team accessed sequencing data within the MD Anderson database and determined variants of unknown significance, upon physician request. The mutations identified and tested within the functional genomics platform for all variants referenced within the paper are available at <https://ibl.mdanderson.org/fasmic/#/>. The accession number is FASMIC00230421.

## CODE AVAILABILITY

Code is available for generation of bioinformatic tool-predicted oncogenicity values (Supplementary Fig. 2) at <https://github.com/KChen-lab/Data-Analysis-of-Variant-Functional-Effects>.

Received: 7 December 2022; Accepted: 3 July 2023;

Published online: 15 July 2023

## REFERENCES

- Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* <https://doi.org/10.1038/nm.4333> (2017).
- Li, M. M. et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A Joint Consensus Recommendation Of The Association For Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn.* **19**, 4–23 (2017).
- U.S. Food and Drug Administration. CDRH'S Approach to Tumor Profiling next Generation Sequencing Tests [Fact sheet]. <https://www.fda.gov/media/109050/download> (2017).
- Mateo, J. et al. A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). *Ann. Oncol.* **29**, 1895–1902 (2018).
- Andre, F. et al. Genomics to select treatment for patients with metastatic breast cancer. *Nature* <https://doi.org/10.1038/s41586-022-05068-3> (2022).
- Sukhai, M. A. et al. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. *Genet. Med.* **18**, 128–136 (2016).
- Meric-Bernstam, F. et al. A decision support framework for genomically informed investigational cancer therapy. *J. Natl. Cancer Inst.* <https://doi.org/10.1093/jnci/djv098> (2015).
- Andre, F. et al. Prioritizing targets for precision cancer medicine. *Ann. Oncol.* **25**, 2295–2303 (2014).
- Van Allen, E. M. et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688 (2014).
- Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00011> (2017).
- Wagner, A. H. et al. A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer. *Nat. Genet.* **52**, 448–457 (2020).
- Peng, R. et al. From somatic variants toward precision oncology: an investigation of reporting practice for next-generation sequencing-based circulating tumor DNA analysis. *Oncologist* **25**, 218–228 (2020).
- Servant, N. et al. Bioinformatics for precision medicine in oncology: principles and application to the SHIVA clinical trial. *Front. Genet.* **5**, 152 (2014).
- Xu, Q. et al. OncoPDSS: an evidence-based clinical decision support system for oncology pharmacotherapy at the individual level. *BMC Cancer* **20**, 740 (2020).
- Dumbrava, E. I. & Meric-Bernstam, F. Personalized cancer therapy-leveraging a knowledge base for clinical decision-making. *Cold Spring Harb. Mol. Case Study* <https://doi.org/10.1101/mcs.a001578> (2018).
- Ghazani, A. A. et al. Assigning clinical meaning to somatic and germ-line whole-exome sequencing data in a prospective cancer precision medicine study. *Genet. Med.* **19**, 787–795 (2017).
- Mosele, F. et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. *Ann. Oncol.* **31**, 1491–1505 (2020).
- Good, B. M., Ainscough, B. J., McMichael, J. F., Su, A. I. & Griffith, O. L. Organizing knowledge to enable personalization of medicine in cancer. *Genome Biol* **15**, 438 (2014).
- Ritter, D. I. et al. Somatic cancer variant curation and harmonization through consensus minimum variant level data. *Genome Med.* **8**, 117 (2016).
- Leichsenring, J. et al. Variant classification in precision oncology. *Int. J. Cancer* **145**, 2996–3010 (2019).
- Zeng, J. et al. Operationalization of next-generation sequencing and decision support for precision oncology. *JCO Clin. Cancer Inform.* **3**, 1–12 (2019).
- Johnson, A. et al. Clinical use of precision oncology decision support. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00036> (2017).
- Nakamura, I. T. et al. Comprehensive functional evaluation of variants of fibroblast growth factor receptor genes in cancer. *NPJ Precis. Oncol.* **5**, 66 (2021).
- Jia, X. et al. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. *Am. J. Hum. Genet.* **108**, 163–175 (2021).
- Findlay, G. M. et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
- Mighell, T. L., Evans-Dutson, S. & O'Roak, B. J. A saturation mutagenesis approach to understanding PTEN lipid phosphatase activity and genotype-phenotype relationships. *Am. J. Hum. Genet.* **102**, 943–955 (2018).
- Kohsaka, S. et al. A method of high-throughput functional evaluation of EGFR gene variants of unknown significance in cancer. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aan6566> (2017).
- Woods, N. T. et al. Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *NPJ Genom. Med.* <https://doi.org/10.1038/npjgenmed.2016.1> (2016).
- Boonen, R., Vreeswijk, M. P. G. & van Attikum, H. Functional characterization of PALB2 variants of uncertain significance: toward cancer risk and therapy response prediction. *Front. Mol. Biosci.* **7**, 169 (2020).
- Zimmerman, L. et al. A novel system for functional determination of variants of uncertain significance using deep convolutional neural networks. *Sci. Rep.* **10**, 4192 (2020).
- Ng, P. K. et al. Systematic functional annotation of somatic mutations in cancer. *Cancer Cell* **33**, 450–462.e410 (2018).
- Johnson, A. et al. The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. *Drug Discov Today* **20**, 1433–1438 (2015).
- Chakravarty, D. et al. Somatic genomic testing in patients with metastatic or advanced cancer: ASCO provisional clinical opinion. *J. Clin. Oncol.* **40**, 1231–1258 (2022).
- Kurnit, K. C. et al. Precision oncology decision support: current approaches and strategies for the future. *Clin. Cancer Res.* **24**, 2719–2731 (2018).
- Markham, A. Erdafitinib: first global approval. *Drugs* **79**, 1017–1021 (2019).
- Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
- Niu, B. et al. Protein-structure-guided discovery of functional mutations across 19 cancer types. *Nat. Genet.* **48**, 827–837 (2016).
- Tokheim, C. et al. Exome-scale discovery of hotspot mutation regions in human cancer using 3D protein structure. *Cancer Res.* **76**, 3719–3731 (2016).
- Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
- Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
- Rogers, M. F. et al. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**, 511–513 (2018).
- Kaminker, J. S., Zhang, Y., Watanabe, C. & Zhang, Z. CanPredict: a computational tool for predicting cancer-associated missense mutations. *Nucleic Acids Res.* **35**, W595–W598 (2007).
- Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
- Bromberg, Y. & Rost, B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* **35**, 3823–3835 (2007).
- van der Velde, K. J. et al. GAVIN: Gene-Aware Variant Interpretation for medical sequencing. *Genome Biol.* **18**, 6 (2017).
- Frazer, J. et al. Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021).
- Tamborero, D. et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10**, 25 (2018).
- Muinos, F., Martinez-Jimenez, F., Pich, O., Gonzalez-Perez, A. & Lopez-Bigas, N. In silico saturation mutagenesis of cancer genes. *Nature* **596**, 428–432 (2021).
- Douville, C. et al. Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). *Hum. Mutat.* **37**, 28–35 (2016).
- Rogers, M. F., Gaunt, T. R. & Campbell, C. CScape-somatic: distinguishing driver and passenger point mutations in the cancer genome. *Bioinformatics* **36**, 3637–3644 (2020).



51. Carter, H. et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res.* **69**, 6660–6667 (2009).
52. Mao, Y. et al. CanDrA: cancer-specific driver missense mutation annotation with optimized features. *PLoS ONE* **8**, e77945 (2013).
53. Li, Y. et al. e-MutPath: computational modeling reveals the functional landscape of genetic mutations rewiring interactome networks. *Nucleic Acids Res.* **49**, e2 (2021).
54. Miosge, L. A. et al. Comparison of predicted and actual consequences of missense mutations. *Proc. Natl Acad. Sci. USA* **112**, E5189–E5198 (2015).
55. Suybeng, V., Koepfel, F., Harle, A. & Rouleau, E. Comparison of pathogenicity prediction tools on somatic variants. *J. Mol. Diagn.* **22**, 1383–1392 (2020).
56. Tokheim, C. & Karchin, R. CHASMPplus reveals the scope of somatic missense mutations driving human cancers. *Cell Syst.* **9**, 9–23.e28 (2019).
57. Martinez-Saez, O. et al. Frequency and spectrum of PIK3CA somatic mutations in breast cancer. *Breast Cancer Res.* **22**, 45 (2020).
58. UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
59. Trevino, V. HotSpotAnnotations—a database for hotspot mutations and annotations in cancer. *Database* <https://doi.org/10.1093/database/baaa025> (2020).
60. Modest, D. P. et al. KRAS allele-specific activity of sunitinib in an isogenic disease model of colorectal cancer. *J. Cancer Res. Clin. Oncol.* **139**, 953–961 (2013).
61. Hunter, J. C. et al. Biochemical and structural analysis of common cancer-associated KRAS mutations. *Mol. Cancer Resm* **13**, 1325–1335 (2015).
62. Cespedes, M. V. et al. K-ras Asp12 mutant neither interacts with Raf, nor signals through Erk and is less tumorigenic than K-ras Val12. *Carcinogenesis* **27**, 2190–2200 (2006).
63. Feldser, D. M. & Kern, S. E. Oncogenic levels of mitogen-activated protein kinase (MAPK) signaling of the dinucleotide KRAS2 mutations G12F and GG12-13VC. *Hum. Mutat.* **18**, 357 (2001).
64. Acunzo, M. et al. Selective targeting of point-mutated KRAS through artificial microRNAs. *Proc. Natl Acad. Sci. USA* **114**, E4203–E4212 (2017).
65. Berger, A. H. et al. High-throughput phenotyping of lung cancer somatic mutations. *Cancer Cell* **30**, 214–228 (2016).
66. Coulier, F., Batoz, M., Marics, I., de Lapeyriere, O. & Birnbaum, D. Putative structure of the FGF6 gene product and role of the signal peptide. *Oncogene* **6**, 1437–1444 (1991).
67. Ropiquet, F., Giri, D., Kwabi-Addo, B., Mansukhani, A. & Ittmann, M. Increased expression of fibroblast growth factor 6 in human prostatic intraepithelial neoplasia and prostate cancer. *Cancer Res.* **60**, 4245–4250 (2000).
68. Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892–897 (2002).
69. Lindstrom, E., Shimokawa, T., Toftgard, R. & Zaphiropoulos, P. G. PTC mutations: distribution and analyses. *Hum. Mutat.* **27**, 215–219 (2006).
70. Mooz, J. et al. Dimerization of the kinase ARAF promotes MAPK pathway activation and cell migration. *Sci. Signal.* **7**, ra73 (2014).
71. Imielinski, M. et al. Oncogenic and sorafenib-sensitive ARAF mutations in lung adenocarcinoma. *J. Clin. Invest.* **124**, 1582–1586 (2014).
72. Pagel, K. A. et al. Integrated informatics analysis of cancer-related variants. *JCO Clin. Cancer Inform.* **4**, 310–317 (2020).
73. Carter, H., Douville, C., Stenson, P. D., Cooper, D. N. & Karchin, R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **14**, S3 (2013).
74. Zhou, W. et al. TransVar: a multilevel variant annotator for precision genomics. *Nat. Methods* **12**, 1002–1003 (2015).
75. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* <https://doi.org/10.1093/gigascience/giab008> (2021).

## ACKNOWLEDGEMENTS

This work was supported by the Sheikh Khalifa Bin Zayed Al Nahyan Institute for Personalized Cancer Therapy, the Cancer Prevention and Research Institute of Texas (RP150535), the Center for Clinical and Translational Science Award (1UL1TR003167), and the MD Anderson Cancer Center Support Grant (2P30CA016672). MD Anderson receives licensing fees for our Precision Oncology Decision Support (PODS) database from Philips Healthcare, which support continued development of the system at MD Anderson. The MD Anderson Cancer Center Precision Oncology Decision Support provides a Decision Support Service for genomic annotation for clinical trials.

## AUTHOR CONTRIBUTIONS

F.M.B. and K.R.M.S. oversaw the project in its entirety. P.N. and B.A. performed the functional genomics experiments. G.B.M. oversaw the functional genomics platform. J.C., T.C., M.K., and A.J. interpreted, tracked, and ingested functional genomics data

into the PODS knowledgebase. J.Z. and A.J. oversaw the requirements for the development of the PODS applications housing the functional genomics and PODS annotation data. A.J., V.H., T.V., F.S., M.K., and S.K. contributed to PODS classification of variants. X.J. conducted the Fisher's exact test for Fig. 4a. Y.W. applied the informatic tool prediction pipelines and conducted the Fisher's exact tests for Fig. 4b and Supplementary Fig. 2. J.Z. developed software to computationally provide a PODS Unknown or Potentially actionability call. T.Y., J.R., and F.M.B. provided clinical oversight for PODS variant classification. A.J. analyzed the data presented, generated the figures and tables, and primarily wrote the manuscript. All authors contributed to editing of the manuscript.

## COMPETING INTERESTS

MD Anderson receives licensing fees for our Precision Oncology Decision Support (PODS) database from Philips Healthcare, which support continued development of the system at MD Anderson. J.R. reports serving on the Advisory Board for Peptomyc, Kelun Pharmaceuticals/Klus Pharma, Ellipses Pharma, Molecular Partners, IONCTURA; Research Funding to his institution from Blueprint Medicines, Black Diamond Therapeutics, Merck Sharp & Dohme, Hummingbird, Yingli, Vall d'Hebron Institute of Oncology/Cancer Core Europe; Clinical Research to his institution from Novartis, Spectrum Pharmaceuticals, Symphogen, BioAlta, Pfizer, GenMab, CytomX, Kelun-Biotech, Takeda-Millennium, GalxoSmithKline, Taiho, Roche Pharmaceuticals, Hummingbird, Yingli, Bicycle Therapeutics, Merus, Curis, Bayer, AadiBioscience, Nuvation, ForeBio, BioMed Valley Discoveries, Loxo Oncology, Hutchinson MediPharma, Cellestia, Deciphera, Ideaya, Amgen, Tango Therapeutics, Mirati Linnaeus Therapeutics; Travel Reimbursement from European Society for Medical Oncology and Other from Vall d'Hebron Institute of Oncology/Ministero De Empleo Y Seguridad Social, Chinese University of Hong Kong, Boxer Capital, LLC, Tang Advisors, LLC. G.B.M. reports SAB/Consultant fees from Amphista, Astex, AstraZeneca, BlueDot, Chrysalis Biotechnology, Ellipses Pharma, GSK, ImmunoMET, Infinity, Ionis, Leapfrog Bio, Lilly, Medacorp, Nanostring, Nuvectis, PDX Pharmaceuticals, Qureator, Roche, Signalchem Lifesciences, Tarveda, Turbine, Zentalis Pharmaceuticals; Stock/Options/Financials from BlueDot, Catena Pharmaceuticals, ImmunoMet, Nuvectis, SignalChem, Tarveda, Turbine; Licensed Technology from HRD assay to Myriad Genetics, DSP patents with Nanostring; Sponsored research from AstraZeneca. T.A.Y. reports Employment at University of Texas MD Anderson Cancer Center, where I am Medical Director of the Institute for Applied Cancer Science, which has a commercial interest in DDR and other inhibitors (IACS30380/ARTO380 was licensed to Artios); Grant/Research support to his institution from Acrivon, Artios, AstraZeneca, Bayer, Beigene, BioNTech, Blueprint, BMS, Clovis, Constellation, Cytel, Eli Lilly, EMD Serono, Forbuis, F-Star, GlaxoSmithKline, Genentech, Haihe, ImmuneSensor, Ionis, Ipsen, Jounce, Karyopharm, KSQ, Kyowa, Merck, Mirati, Novartis, Pfizer, Ribon Therapeutics, Regeneron, Repare, Rubius, Sanofi, Scholar Rock, Seattle Genetics, Tesaro, Vivace and Zenithl Consultancy fees from AbbVie, AstraZeneca, Acrivon, Adagene, Almac, Aduro, Amphista, Artios, Athena, Atrin, Avoro, Axiom, Baptist Health Systems, Bayer, Beigene, Boxer, Bristol Myers Squibb, C4 Therapeutics, Calithera, Cancer Research UK, Clovis, Cybrexa, Diffusion, EMD Serono, F-Star, Genmab, Glenmark, GLG, Globe Life Sciences, GSK, Guidepoint, Idience, Ignyta, I-Mab, ImmuneSensor, Institut Gustave Roussy, Intellisphere, Jansen, Kyn, MEI pharma, Mereo, Merck, Natera, Nexys, Novocure, OHSU, OncoSec, Ono Pharma, Pegascy, PER, Pfizer, Piper-Sandler, Prolynx, Repare, resTORbio, Roche, Schrodinger, Theragnostics, Varian, Versant, Vibliome, Xinthera, Zai Labs and ZielBio; he also holds stock in Seagen. F.M-B. reports Consulting fees from AbbVie, Aduro BioTech Inc., Alkermes, AstraZeneca, Daiichi Sankyo Co. Ltd., DebioPharm, Ecor1 Capital, eFFECTOR Therapeutics, F. Hoffman-La Roche Ltd., GT Apeiron, Genentech Inc., Harbinger Health, IBM Watson, Infinity Pharmaceuticals, Jackson Laboratory, Kolon Life Science, Lengo Therapeutics, Menarini Group, OrigiMed, PACT Pharma, Parexel International, Pfizer Inc., Protai Bio Ltd, Samsung Bioepis, Seattle Genetics Inc., Tallac Therapeutics, Tyra Biosciences, Xencor, Zymeworks; has served on the Advisory Committee for Black Diamond, Biovica, Eisai, FogPharma, Immunomedics, Inflection Biosciences, Karyopharm Therapeutics, Loxo Oncology, Mersana Therapeutics, OnCusp Therapeutics, Puma Biotechnology Inc., Seattle Genetics, Sanofi, Silverback Therapeutics, Spectrum Pharmaceuticals, Zentalis; has received Sponsored Research (to the institution) from Aileron Therapeutics, Inc. AstraZeneca, Bayer Healthcare Pharmaceutical, Calithera Biosciences Inc., Curis Inc., CytomX Therapeutics Inc., Daiichi Sankyo Co. Ltd., Debiopharm International, eFFECTOR Therapeutics, Genentech Inc., Guardant Health Inc., Klus Pharma, Takeda Pharmaceutical, Novartis, Puma Biotechnology Inc., Taiho Pharmaceutical Co.; has received Honoraria (speaking engagement) from Chugai Biopharmaceuticals and Travel reimbursement from European Organisation for Research and Treatment of Cancer (EORTC) and the European Society for Medical Oncology (ESMO). The remaining authors declare no competing interests.

**ADDITIONAL INFORMATION**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-023-00420-w>.

**Correspondence** and requests for materials should be addressed to Funda Meric-Bernstam.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023