BBA - Reviews on Cancer 1876 (2021) 188614

Contents lists available at ScienceDirect

BBA - Reviews on Cancer

journal homepage: www.elsevier.com/locate/bbacan

Review

Protein domain-based approaches for the identification and prioritization of therapeutically actionable cancer variants

Elisabetta Grillo [a,*], Cosetta Ravelli [a], Michela Corsini [a], Luca Zammataro [b], Stefania Mitola [a,*]

[a] Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[b] Division of Artificial Intelligence Systems for Immunoinformatics, Kiromic BioPharma, Inc., Houston, USA

ARTICLE INFO

Keywords:
Protein domain
Somatic mutation
Cancer variant
Low frequency mutation

ABSTRACT

The tremendous number of cancer variants that can be detected by NGS analyses has required the development of computational approaches to prioritize mutations on the basis of their biological and clinical significance. Standard strategies take a gene-centric approach to the problem, allowing exclusively the identification of highly frequent variants. On the contrary, protein domain (PD)-based approaches allow to identify functionally relevant low frequency variants by searching for mutations that recur on analogous residues across homologous proteins (i.e. containing the same PD). Such approaches enable to transfer information about the effects and druggability from one known mutation to unknown ones. Here we describe how PD-based strategies work, and discuss how they could be exploited for mutation prioritization.

The principle that mutations clustered on specific residues of PDs have the same functional consequences and are therapeutically actionable in a similar manner could help the choice of patient-specific targeted drugs, eventually improving the management of cancer patients.

1. Introduction

Precision oncology refers to the possibility of using therapeutics that are specifically tailored to the genetic changes in each person's cancer. There is widespread enthusiasm for precision oncology, also fueled by the increasing number of cancer-associated genetic changes that can be detected in large cohorts of cancer specimens [e.g., The Cancer Genome Atlas (TCGA) [1], the Cancer Genome Project (CGP) [2], the International Cancer Genome Consortium [3] and the Catalogue of Somatic Mutations in Cancer (COSMIC) [4] through next-generation sequencing (NGS) technologies and by the increasing number of targeted agents becoming available. Understanding the complete landscape of mutations that drive cancer remains an open challenge for precision oncology and it could help in revealing novel tumor-specific therapeutic vulnerabilities [5]. Many research efforts have been spent on accurately characterizing all these molecular alterations. The development of novel computational algorithms to search, analyze and prioritize cancer-associated mutations have led to extraordinary advances [6–12]. These efforts have focused on highly frequent variants, disregarding instead rare and low-frequency mutations. Thus, clinical benefits have been obtained only for a restricted group of cancers, including the frequent BCR-ABL1-fusion-positive chronic myeloid leukemia and BRAF mutated melanoma [13,14], while most tumors have profited much less from precision oncology. This manifests the urgent need for the identification of clinically relevant mutations raising the question of how to identify and characterize clinically relevant rare or low-frequency driver mutations [15]. An accurate characterization of patient-specific molecular alterations could accelerate the process of drug selection, bringing significant clinical benefits. However, the study of each of these variants is time-consuming and extremely expensive, not affordable and applicable in routine clinical activities in economically developed countries and unthinkable in most of other countries.

In this article, we discuss current approaches useful to overcome these limitations that we believe could maximize the benefits of personalized medicine. We will focus on the analysis of cancer mutations at the level of protein domains (PDs) to identify molecular predictors for the stratification of patients and drug selection.

2. Gene mutations in cancer

Cancer develops following the acquisition and accumulation of gene mutations due to the exposure to chemicals or radiations, hormones,

* Corresponding authors at: Department of Molecular and Translational Medicine, University of Brescia, Via Branze 39, 25123 Brescia, Italy.
E-mail addresses: elisabetta.grillo@unibs.it (E. Grillo), stefania.mitola@unibs.it (S. Mitola).

https://doi.org/10.1016/j.bbcan.2021.188614
Received 22 July 2021; Received in revised form 11 August 2021; Accepted 11 August 2021
Available online 14 August 2021

aging or other environmental factors. Over time, a number of mutations may accumulate in a single cell, promoting abnormal proliferation and eventually tumorigenesis [16]. In addition to somatic mutations that are acquired during life, inherited germline mutations represent one of the major risk factors in cancer [17]. Here we will focus on somatic cancer mutations. However, the domain-based approaches we discuss in this review may also be applied to germline mutations. Many public databases provide still very little information concerning these variants. A recent update of cBioPortal improved visualization of germline mutations, which anyway remain the minority of all mutations found in this database.

2.1. *Driver mutations*

Those mutations that confer a growth advantage to cancer cells and are able to initiate or trigger the progression of cancer are called ‘driver’ mutations, while all the others are termed ‘passenger’ variants [18]. Distinguishing between the two classes remains a significant challenge. Traditionally, driver mutations were identified based on their frequency in cancer patients. However, various examples of allosteric driver mutations or rare driver mutations have been provided and could be exploited for precision medicine [15,19].

The introduction of NGS technologies have allowed a fast, cheap and accurate detection of whole-genome pan-cancer mutations and have increased the power of discriminating between driver and passenger mutations, allowing to pinpoint the causative and/or therapeutically actionable variants. For example, MH Bailey and colleagues performed a pan-cancer and pan-software analysis and identified in silico more than 3400 missense driver mutations. More than 60% of these mutations were experimentally validated as drivers or putative clinically actionable mutations [20].

2.2. *Frequent or rare mutations*

The most characterized variants occur at high-frequency in cancer patients and include, among others, the hotspot mutation V600E of B-Raf in melanoma [21] or APC mutations in colon carcinoma [22]. More recently, large collections of cancer genomes and pan-cancer analyses have led to the identification of similar tumorigenic mutations in different cancers including mutations of TP53 (found in 42% of samples in a pan cancer analysis) PIK3CA, MLL2/3/4, KRAS, NRAS and EGFR [23].

On the other hand, rare mutations, defined by several authors as those occurring in less than 1% of cancer patients [15,24], dominate the landscape of cancer-associated alterations. These can drive tumorigenesis and can be therapeutically targeted [15,19]. Remarkably, tumors are heterogeneous and dynamic organs in which a mutation occurs only in a sub-clone in a specific moment of tumor progression. Rare and sub-clonal mutations remained unexplored until the introduction of high sensitivity NGS technologies [25]. Although these rare mutations may be responsible for drug response, tumor relapse or metastasis, they were disregarded for their expensive translational potential. However, understanding their effects on tumor behavior and on the therapeutic response is of great importance for precision oncology. Various bioinformatics and bio-statistical approaches have been proposed for the identification, validation and the analysis of the druggability of rare variants [19,26–29].

3. Protein domains: functional regions

Proteins generally contain one or more functional regions in their sequence, which are commonly termed ‘domains’. In 1973, DB Wetlaufer defined for the first time PDs as stable units of a protein structure that could fold autonomously [30]. PDs retain their cognate biochemical function and 3D structure even when isolated from the protein context. PDs are evolutionary conserved in terms of amino acid sequence, 3D structure and function. The presence of different combinations of domains in different proteins gives rise to the diverse functional protein repertoire found in nature [31,32]. Identifying which domains are present in a given protein provides insight into the function of that protein.

Various databases have been created for the classification, analysis and identification of PDs. These allow to transfer information from characterized sequences to uncharacterized ones. The Structural Classification of Proteins (SCOP) database, created by manual inspection, classifies the PDs of known proteins in tree structures considering evolutionary relationships (family and superfamily) and those that arise from physics and chemistry of proteins (class). The peculiarity of SCOP is to exploit the evolutionary relationships among PDs which imply functional similarity better than simple physical relationships [33,34]. SCOP was later updated to SCOP2 in which the structural and evolutionary relationships are separated, allowing the classification of the homologous proteins into different folds and structural classes while keeping them in the same evolutionary family and superfamily [35].

More recently, the InterPro consortium (/http://www.ebi.ac.uk/interpro/about/consortium/) collected 13 different databases for protein analysis for PD identification and classification including CATH, CDD, HAMAP, PANTHER, Prosite, Pfam and others [36]. Most of their algorithms are based on the Hidden Markov Model (HMM). The profile HMM is trained on a small representative set of aligned sequences that are known to belong to the family. This model is then used to search all homologous sequences against a large sequence database as UniProtKB to build those that we term ‘protein families’, consisting in groups of proteins that include in their sequence a given domain. On these bases the same protein containing more than one PD will belong to the corresponding distinct protein families. Remarkably, some databases, such as Pfam [37], classify PDs based on their linear amino acid sequences, while others, including CATH [38], exploit three-dimensional structures downloaded from the protein data bank (PDB) to catalogue PDs. In the following chapters, PDs will be named according to their Pfam classification.

4. Domain-based approaches for the identification of mutation hotspots

Over the last few years, numerous ‘gene-based’ methods for the identification of putative oncogenic variants have been developed. By analyzing one single gene at a time they take into account a variety of features, including the evolutionary story of the mutated position, the recurrence of the mutation across different cancer types or its expected effects on protein structure, stability and function [39–42]. However, gene-based approaches have the potential to mischaracterize mutations [43]. Also, different mutations in the same gene can trigger different functional consequences. L. Shen and colleagues classified 33 mutated genes that can act both as oncogenes and tumor suppressors [44]. This may be due to different mutations within the same PD or mutations in different PDs of the same protein. To overcome these limitations alternative methods, including tools to analyze cancer mutations at PD level, have been proposed (Fig. 1).

PD-based approaches search for mutations that cluster on specific amino acid positions (called ‘mutation hotspots’) within a given PD enabling the transfer of functional and clinical information from known variants to uncharacterized mutations. This defines what MG Kann called ‘domain landscape’ [45]. Domain-based methods for mutation analysis rely on the hypothesis that variants clustered on analogous residues within a PD across different proteins have a high probability to elicit similar functional consequences and be therapeutically actionable in a similar manner.

PD-based strategies aggregate mutational data from different proteins containing a given domain and generate a unique mutational profile. The general workflow of PD-based mutational approaches (summarized in Fig. 2) entails 4 steps:

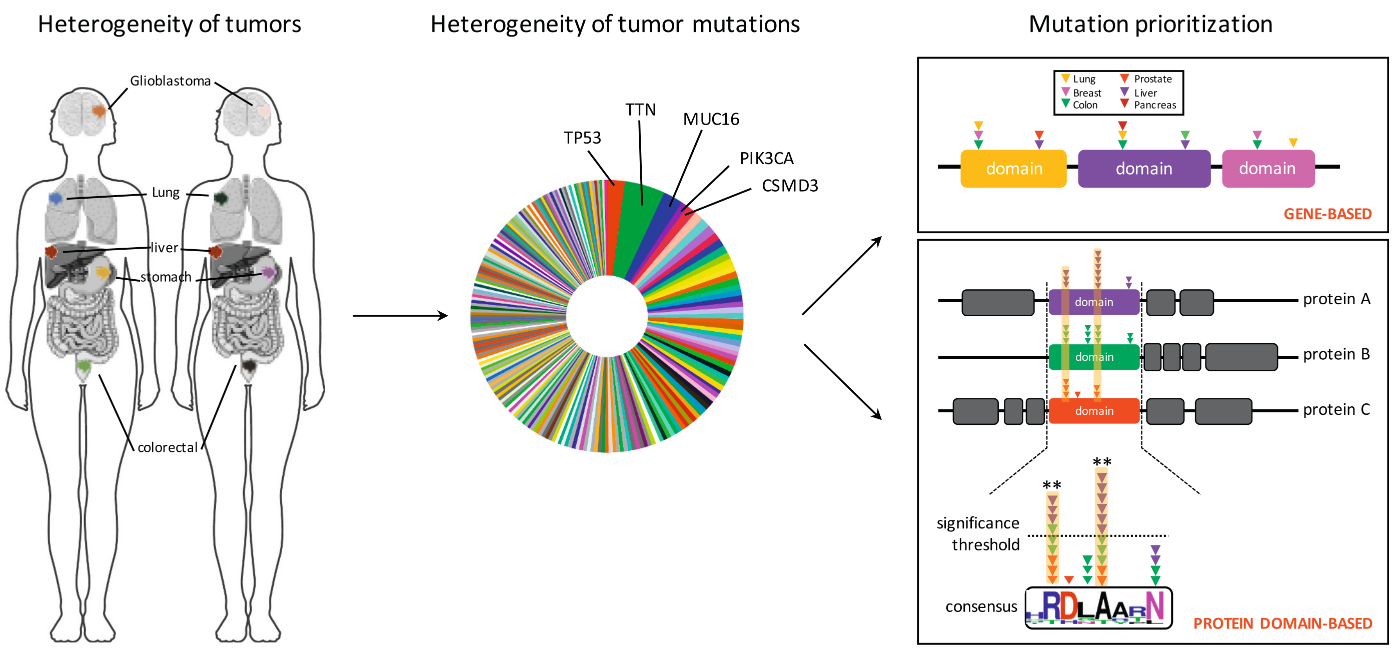

Fig. 1. Mutation prioritization for the discovery of new functionally relevant cancer variants. Tumors are heterogeneous entities both in space (i.e. the organ involved) and in time (i.e. the stage of the tumor). In addition, inter-individual variation exists. The tremendous number of cancer variants that can be detected by NGS requires the development of computational approaches to prioritize mutations in order to select the ones with a high probability to be oncogenic and/or therapeutically actionable. Standard gene-based approaches have exploited the recurrence of mutations in one gene across different cancer types to pinpoint potential new driver variants. On the other hand, PD-based approaches allow the identification of hotspots of mutations clustered on analogous residues across homologous proteins (i.e. containing the same PD) across cancer types. Such approaches increase the power of the analyses and enable to transfer functional information from one known mutation to unknown ones, thus inferring the functional consequences of a given mutation.

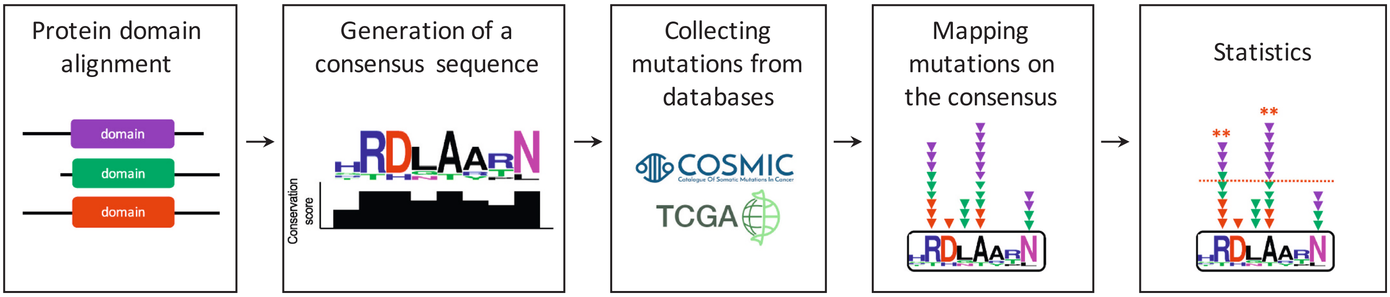
Protein domain alignment
domain
domain
domain
Generation of a consensus sequence
RDLAARN
Conservation score
Collecting mutations from databases
COSMIC
Catalogue Of Somatic Mutations In Cancer
TCGA
Mapping mutations on the consensus
RDLAARN
Statistics
**
**
RDLAARN

Fig. 2. Workflow of PD-based approaches for the identification of novel cancer-associated variants.

A. PD alignment: the first step is the generation of a multiple sequence alignment of the PDs of all proteins of interest. This is usually achieved through alignment tools such as ClutalO.
B. Creation of a consensus sequence: a sequence including the most represented amino acid found at every position is created starting from the multiple sequence alignment. Some tools also calculate a per-position conservation score, which highlights the functionally important residues within a given PD.
C. Collection and mapping of mutations: somatic mutations of all target proteins are retrieved from public or custom databases and mapped per-position along the consensus sequence. This step allows to cluster all analogous mutations on a single position of the aligned domain.
D. Statistics: statistical models are used to calculate the statistical significance of clustered mutations considering: i) the conservation score of each position; ii) a significance threshold of background mutation frequency.

This general workflow can be integrated or complemented with tools to predict the functional impact of the identified mutations [46], resources to map the mutations on the 3D structure of a given protein [47–51], software that align the 3D structures instead of linear amino acid sequences [52], methods to visualize the position of the mutations in single proteins or the analysis of co-occurrence and mutual exclusivity of mutations [53]. Altogether this workflow defines a ‘domain-based mutation prioritization’. It enables navigating cancer mutations and prioritizing them on the bases of the putative functional impact, inferred from their position within a given domain or from the correspondence with other known oncogenic variants. Also, the aggregation of mutations from several genes increases the statistical power of rare and functional variants (Box 1).

5. PD-based mutation analysis: from genes to drug targets

In 2010, MG Kann’s team released the first bioinformatic resource for PD-based mutational analysis, called domain mapping of disease mutations (DMDM) database [54]. DMDM enables an aggregated view of all human coding disease-related mutations for each PD. It retrieves mutation data from OMIM, Swiss-Prot, and single nucleotide polymorphisms (SNPs) from the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/). By applying DMDM to the analysis of the landscape of

Box 1

Domain-based analyses of cancer mutations enable:

1) a rational mutation prioritization. Mutation hotspots have a high probability to elicit functional consequences, thus helping to select functionally significant mutations among negligible alterations
2) to pinpoint rare variants that would remain disregarded using gene-based strategies
3) to infer the possible mechanism of action of an unknown mutation on the basis of its position within the domain
4) to find correspondence between novel unknown variants and known oncogenic or druggable mutations, thus inferring the possible functional consequences of a given mutation

mutations in colon adenocarcinoma tumor and breast invasive carcinoma, the authors extracted the most significantly mutated PDs and identified various mutation hotspots. Among those, some hotspots aggregate mutations from single genes, while others, including the ones occurring in the CENP-B_N domain, aggregate mutations from numerous genes. By doing so, they demonstrated that through PD-based approaches also low frequency variants reach the statistical significance threshold [45,55]. Interestingly, DMDM compares the cancer-related variants with variants occurring in other human diseases. This peculiar feature of DMDM increases the possibilities to transfer functional information from one known variant to unknown ones [54,56]. In the same period, the group of G. Cavet released a similar interactive web application for browsing mutation clusters among germline and somatic mutations found in cancer and other diseases [57].

In 2015, F Yang and colleagues demonstrated that in the same cancer sample a given hotspot mutation was not present in more than one protein belonging to the same family. Thus, the mutations aggregated in the same hotspot are mutually exclusive [58]. This strongly indicates that analogous mutations act in the same pathway via a parallel mechanism.

More recently, the MutationAligner web resource was developed (http://mutationaligner.org) [59,60]. This user-friendly interactive interface enables visualizing, navigating and analyzing somatic mutation hotspots identified in PDs from TCGA variant data processed by cBioPortal up to spring 2015 across genes and 22 tumor types. Despite this, MutationAligner does not allow to select subgroups of target proteins or select the type of cancer to refine the analysis. Using MutationAligner, ML Miller et al. identified 14 PDs significantly enriched for missense mutations, including the tyrosine kinase (PK_Tyr_Ser-Thr), P53, Ras and Cadherin domains [59]. This is not surprising, as all the identified domains exert biological roles in cell proliferation, differentiation and metastasization. Miller's analyses also highlighted the correspondence between some well-known cancer driver mutations such as B-Raf^{V600E} and KIT$^{D816V/Y}$ with uncharacterized low frequency mutations [59,61]. In a similar way, RA Toledo and colleagues showed that seven mutations of the vascular endothelial growth factor receptor 2 (VEGFR2) lay in mutation hotspots recurring in the PK_Tyr_Ser-Thr domain. Functional analysis supported the tumorigenic effects of substitutions L840F and R1032Q of VEGFR2 [62].

The bioinformatic tool LowMACA (Low frequency Mutations Analysis via Consensus Alignment) was designed for the identification of low frequency variants in pan-cancer analysis [53]. LowMACA is a versatile platform that allows to analyze Pfams, subgroups of proteins within a given Pfam or a custom protein list in all or specific cancer types. LowMACA retrieves cancer somatic mutations by querying cBioportal or other freely available databases. Also, LowMACA allows importing mutations from homemade datasets. Using LowMACA G Melloni et al. analyzed the RAS superfamily and identified new putative driver mutations in Rho, Rab and Rheb subfamilies that would have remained disregarded if single genes were considered [53]. By analyzing a subgroup of tyrosine kinase receptors belonging to the PK_Tyr_Ser-Thr protein family via LowMACA we recently found and characterized the previously unknown R1051Q and D1052N mutations of VEGFR2 as activating mutations [61,63]. Combining LowMACA with MutationAligner we revealed the correspondence between R1051Q and D1052N substitutions of VEGFR2 with the well known oncogenic mutations T599I and V600E of B-Raf [61]. On these bases, all mutations occurring at these positions in any tyrosine/Ser/Thr kinase domain-containing protein have a high probability to elicit pro-oncogenic effects.

Sporadic studies have employed PD-based strategies for the identification of novel driver cancer mutations. For example, it was shown that Pkinase, PK_Tyr_Ser-Thr, Y-Phosphatase and Src-homology 2 domains carry stage-specific mutations in invasive ductal breast cancer [64]. Also, two novel hotspot mutations in the immunoglobulin (Ig)-like domain of FGFR2 were found in colon cancer [58] by using PD-based analyses.

In recent years, to benefit from the accuracy of 3D structure-based PDs, domain-based approaches have been implemented to also include structural data from the PDB database. For example, ZR Moghadam and colleagues took advantage of CATH database to extract structure-based PDs for the identification of cancer-type specific somatic mutations recurring across PDs [52]. By using a CATH-based approach M Ashford et al. calculated the 3D proximity of mutation clusters to catalytic residues, protein-protein interfaces and ligand binding regions. Such analysis filtered out mutations that do not affect protein function (probably passengers) enabling a more accurate detection of functionally relevant mutations [65]. Although these methods are limited to the available 3D structures (only about 75% of proteins have a PDB structure) they have the great advantage to allow the visualization of the 3D spatial information about a given mutation hotspot and to infer from this information its putative functional role.

Beside discovering unknown variants and inferring their role in tumorigenesis, PD-based strategies have the potential of highlighting new therapeutically actionable cancer variants. One can speculate that, for example, all mutations analogous to the substitution V600E of B-Raf oncogene, may exhibit strong sensitivity to Vemurafenib (PLX4032) kinase inhibitor, similarly to mutated B-Raf [66,67]. This hypothesis would definitely increase the number of patients candidate for a given drug with tremendous clinical benefits. The analysis of 3D spatial information may further increase the number of putative actionable variants and target patients by pinpointing those additional mutations that occur in residues close to the ones that confer drug sensitivity [49] (Fig. 3).

6. Conclusions

The possibility to identify and attribute a functional role to mutations that may be causative of cancer is not only an important scientific question for understanding the mechanisms of tumorigenesis, but is also

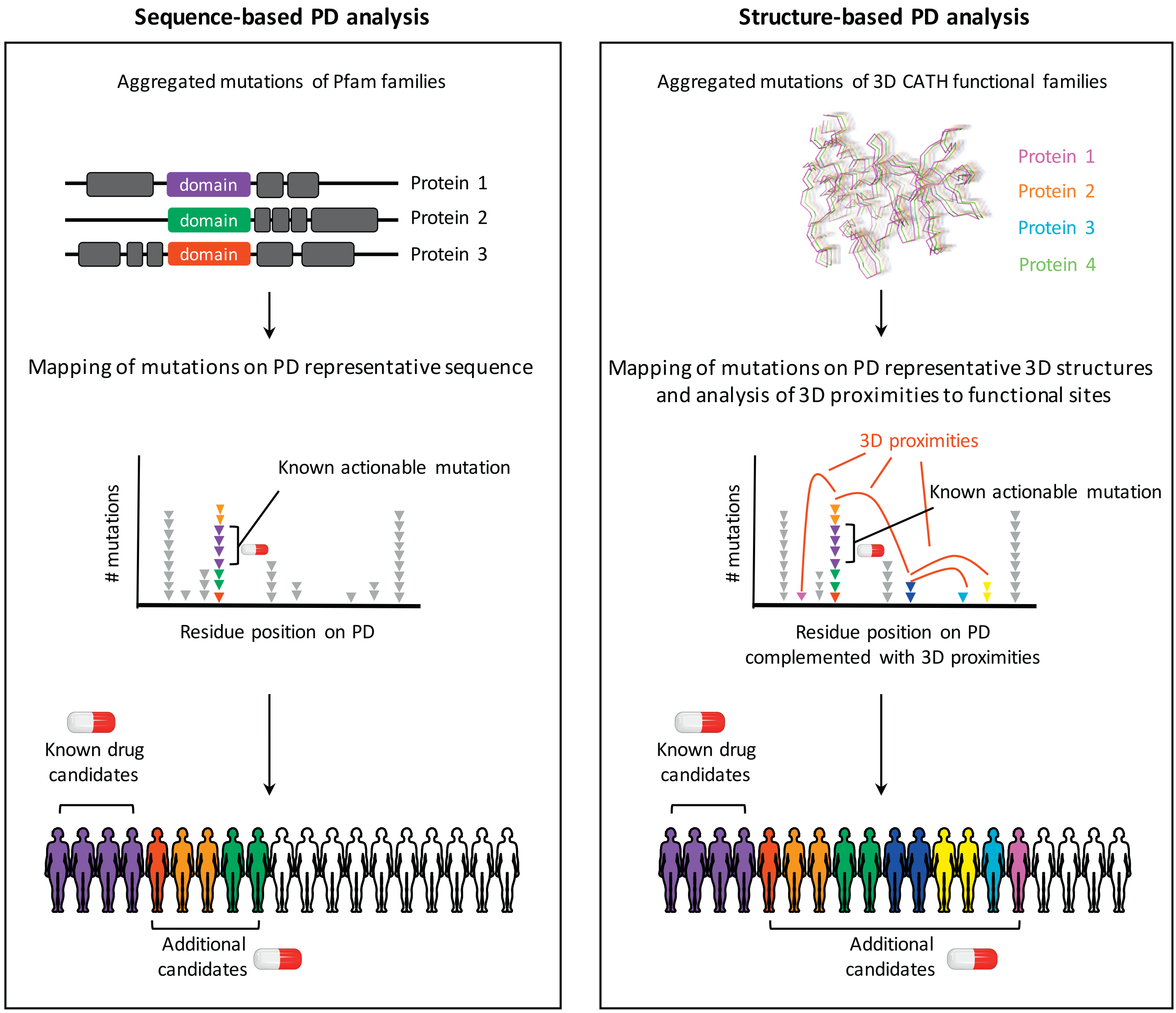
Sequence-based PD analysis
Aggregated mutations of Pfam families
domain
Protein 1
domain
Protein 2
domain
Protein 3
Mapping of mutations on PD representative sequence
# mutations
Known actionable mutation
Residue position on PD
Known drug candidates
Additional candidates
Structure-based PD analysis
Aggregated mutations of 3D CATH functional families
Protein 1
Protein 2
Protein 3
Protein 4
Mapping of mutations on PD representative 3D structures and analysis of 3D proximities to functional sites
3D proximities
# mutations
Known actionable mutation
Residue position on PD complemented with 3D proximities
Known drug candidates
Additional candidates

Fig. 3. PD-based analysis of cancer mutations for the identification of novel druggable variants. The pipelines of sequence- and 3D structure-based approaches for the analysis of therapeutically actionable PD mutations are compared. By considering the 3D spatial distribution of residues and the proximity to known functional sites (e.g. catalytic sites, binding sites, protein interaction interfaces) 3D-based methods have an increased potency in the identification of unknown druggable variants.

essential to the development of personalized cancer therapies.

Here we described the state-of-the-art PD-based approaches for a systematic identification, classification, and investigation of cancer-associated mutations. Altogether, these strategies promise an accurate prioritization of driver mutations in cancer genomes. Despite PD-based approaches enable the transfer of functional information from characterized variants to unknown ones, further efforts are warranted to finally confirm that all mutations clustered on specific residues of PDs do have the same functional consequences. This concept is particularly important for therapeutically actionable mutations. Indeed, PD-based analyses have the potential to accelerate the choice of patient-specific targeted drugs. Also, they set the bases for the development of new cancer drugs. Altogether, these efforts will guarantee a more efficacious application of precision oncology.

Author contributions

E.G, C.R., M.C. drafted the manuscript; E.G. prepared figures; E.G., L. Z., S.M. edited and revised the manuscript.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by grants from AIRC to S.M. (IG17276). E. G. was supported by Fondazione Umberto Veronesi Fellowship.

References

[1] N. Cancer Genome Atlas Research, Comprehensive genomic characterization defines human glioblastoma genes and core pathways, Nature 455 (2008) 1061–1068.

[2] P.A. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, M.R. Stratton, A census of human cancer genes, Nat. Rev. Cancer 4 (2004) 177–183.

[3] C. International Cancer Genome, T.J. Hudson, W. Anderson, A. Artez, A.D. Barker, C. Bell, R.R. Bernabe, M.K. Bhan, F. Calvo, I. Eerola, D.S. Gerhard, A. Guttmacher, M. Guyer, F.M. Hemsley, J.L. Jennings, D. Kerr, P. Klatt, P. Kolar, J. Kusada, D.P. Lane, F. Laplace, L. Youyong, G. Nettekoven, B. Ozenberger, J. Peterson, T.S. Rao, J. Remacle, A.J. Schafer, T. Shibata, M.R. Stratton, J.G. Vockley, K. Watanabe, H. Yang, M.M. Yuen, B.M. Knoppers, M. Bobrow, A. Cambon-Thomsen, L.G. Dressler, S.O. Dyke, Y. Joly, K. Kato, K.L. Kennedy, P. Nicolas, M.J. Parker, E. Rial-Sebbag, C.M. Romeo-Casabona, K.M. Shaw, S. Wallace, G.L. Wiesner, N. Zeps, P. Lichter, A.V. Biankin, C. Chabannon, L. Chin, B. Clement, E. de Alava, F. Degos, M.L. Ferguson, P. Geary, D.N. Hayes, T.J. Hudson, A.L. Johns, A. Kasprzyk, H. Nakagawa, R. Penny, M.A. Piris, R. Sarin, A. Scarpa, T. Shibata, M. van de Vijver, P.A. Futreal, H. Aburatani, M. Bayes, D.D. Botwell, P.J. Campbell, X. Estivill, D.S. Gerhard, S.M. Grimmond, I. Gut, M. Hirst, C. Lopez-Otin, P. Majumder, M. Marra, J.D. McPherson, H. Nakagawa, Z. Ning, X.S. Puente, Y. Ruan, T. Shibata, M.R. Stratton, H.G. Stunnenberg, H. Swerdlow, V.E. Velculescu, R.K. Wilson, H.H. Xue, L. Yang, P.T. Spellman, G.D. Bader, P.C. Boutros, P.J. Campbell, P. Flicek, G. Getz, R. Guigo, G. Guo, D. Haussler, S. Heath, T.J. Hubbard, T. Jiang, S.M. Jones, Q. Li, N. Lopez-Bigas, R. Luo, L. Muthuswamy, B.F. Ouellette, J.V. Pearson, X.S. Puente, V. Quesada, B.J. Raphael, C. Sander, T. Shibata, T.P. Speed, L.D. Stein, J.M. Stuart, J.W. Teague, Y. Totoki, T. Tsunoda, A. Valencia, D.A. Wheeler, H. Wu, S. Zhao, G. Zhou, L.D. Stein, R. Guigo, T.J. Hubbard, Y. Joly, S.M. Jones, A. Kasprzyk, M. Lathrop, N. Lopez-Bigas, B.F. Ouellette, P.T. Spellman, J.W. Teague, G. Thomas, A. Valencia, T. Yoshida, K.L. Kennedy, M. Axton, S.O. Dyke, P.A. Futreal, D.S. Gerhard, C. Gunter, M. Guyer, T.J. Hudson, J.D. McPherson, L.J. Miller, B. Ozenberger, K.M. Shaw, A. Kasprzyk, L.D. Stein, J. Zhang, S.A. Haider, J. Wang, C.K. Yung, A. Cros, Y. Liang, S. Gnaneshan, J. Guberman, J. Hsu, M. Bobrow, D.R. Chalmers, K.W. Hasel, Y. Joly, T.S. Kaan, K.L. Kennedy, B.M. Knoppers, W.W. Lowrance, T. Masui, P. Nicolas, E. Rial-Sebbag, L.L. Rodriguez, C. Vergely, T. Yoshida, S.M. Grimmond, A.V. Biankin, D.D. Bowtell, N. Cloonan, A. de Fazio, J.R. Eshleman, D. Etemadmoghadam, B.B. Gardiner, J.G. Kench, A. Scarpa, R.L. Sutherland, M.A. Tempero, N.J. Waddell, P.J. Wilson, J.D. McPherson, S. Gallinger, M.S. Tsao, P.A. Shaw, G.M. Petersen, D. Mukhopadhyay, L. Chin, R.A. DePinho, S. Thayer, L. Muthuswamy, K. Shazand, T. Beck, M. Sam, L. Timms, V. Ballin, Y. Lu, J. Ji, X. Zhang, F. Chen, X. Hu, G. Zhou, Q. Yang, G. Tian, L. Zhang, X. Xing, X. Li, Z. Zhu, Y. Yu, J. Yu, H. Yang, M. Lathrop, J. Tost, P. Brennan, I. Holcatova, D. Zaridze, A. Brazma, L. Egevard, E. Prokhortchouk, R.E. Banks, M. Uhlen, A. Cambon-Thomsen, J. Viksna, F. Ponten, K. Skryabin, M.R. Stratton, P.A. Futreal, E. Birney, A. Borg, A.L. Borresen-Dale, C. Caldas, J.A. Foekens, S. Martin, J.S. Reis-Filho, A.L. Richardson, C. Sotiriou, H.G. Stunnenberg, G. Thoms, M. van de Vijver, L. van't Veer, F. Calvo, D. Birnbaum, H. Blanche, P. Boucher, S. Boyault, C. Chabannon, I. Gut, J.D. Masson-Jacquemier, M. Lathrop, I. Pauporte, X. Pivot, A. Vincent-Salomon, E. Tabone, C. Theillet, G. Thomas, J. Tost, I. Treilleux, F. Calvo, P. Bioulac-Sage, B. Clement, T. Decaens, F. Degos, D. Franco, I. Gut, M. Gut, S. Heath, M. Lathrop, D. Samuel, G. Thomas, J. Zucman-Rossi, P. Lichter, R. Eils, B. Brors, J.O. Korbel, A. Korshunov, P. Landgraf, H. Lehrach, S. Pfister, B. Radlwimmer, G. Reifenberger, M.D. Taylor, C. von Kalle, P.P. Majumder, R. Sarin, T.S. Rao, M.K. Bhan, A. Scarpa, P. Pederzoli, R.A. Lawlor, M. Delledonne, A. Bardelli, A.V. Biankin, S.M. Grimmond, T. Gress, D. Klimstra, G. Zamboni, T. Shibata, Y. Nakamura, H. Nakagawa, J. Kusada, T. Tsunoda, S. Miyano, H. Aburatani, K. Kato, A. Fujimoto, T. Yoshida, E. Campo, C. Lopez-Otin, X. Estivill, R. Guigo, S. de Sanjose, M.A. Piris, E. Montserrat, M. Gonzalez-Diaz, X.S. Puente, P. Jares, A. Valencia, H. Himmelbauer, V. Quesada, S. Bea, M.R. Stratton, P.A. Futreal, P.J. Campbell, A. Vincent-Salomon, A.L. Richardson, J.S. Reis-Filho, M. van de Vijver, G. Thomas, J.D. Masson-Jacquemier, S. Aparicio, A. Borg, A.L. Borresen-Dale, C. Caldas, J.A. Foekens, H.G. Stunnenberg, L. van't Veer, D.F. Easton, P.T. Spellman, S. Martin, A.D. Barker, L. Chin, F.S. Collins, C.C. Compton, M.L. Ferguson, D.S. Gerhard, G. Getz, C. Gunter, A. Guttmacher, M. Guyer, D.N. Hayes, E.S. Lander, B. Ozenberger, R. Penny, J. Peterson, C. Sander, K.M. Shaw, T.P. Speed, P.T. Spellman, J.G. Vockley, D.A. Wheeler, R.K. Wilson, T.J. Hudson, L. Chin, B.M. Knoppers, E.S. Lander, P. Lichter, L.D. Stein, M.R. Stratton, W. Anderson, A.D. Barker, C. Bell, M. Bobrow, W. Burke, F.S. Collins, C.C. Compton, R.A. DePinho, D.F. Easton, P.A. Futreal, D.S. Gerhard, A.R. Green, M. Guyer, S.R. Hamilton, T.J. Hubbard, O.P. Kallioniemi, K.L. Kennedy, T.J. Ley, E.T. Liu, Y. Lu, P. Majumder, M. Marra, B. Ozenberger, J. Peterson, A.J. Schafer, P.T. Spellman, H.G. Stunnenberg, B.J. Wainwright, R.K. Wilson, H. Yang, International network of cancer genome projects, Nature 464 (2010) 993–998.

[4] S.A. Forbes, G. Tang, N. Bindal, S. Bamford, E. Dawson, C. Cole, C.Y. Kok, M. Jia, R. Ewing, A. Menzies, J.W. Teague, M.R. Stratton, P.A. Futreal, COSMIC (the catalogue of somatic mutations in cancer): a resource to investigate acquired mutations in human cancer, Nucleic Acids Res. 38 (2010) D652–D657.

[5] J.A. Moscow, T. Fojo, R.L. Schilsky, The evidence framework for precision cancer medicine, Nat. Rev. Clin. Oncol. 15 (2018) 183–192.

[6] H. Tsang, K. Addepalli, S.R. Davis, Resources for interpreting variants in precision genomic oncology applications, Front. Oncol. 7 (2017) 214.

[7] C. Tokheim, R. Karchin, CHASMplus reveals the scope of somatic missense mutations driving human cancers, Cell Syst. 9 (2019) 9–23 (e28).

[8] M.A. Sukhai, K.J. Craddock, M. Thomas, A.R. Hansen, T. Zhang, L. Siu, P. Bedard, T.L. Stockley, S. Kamel-Reid, A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer, Genet Med. 18 (2016) 128–136.

[9] N.M. Ioannidis, J.H. Rothstein, V. Pejaver, S. Middha, S.K. McDonnell, S. Baheti, A. Musolf, Q. Li, E. Holzinger, D. Karyadi, L.A. Cannon-Albright, C.C. Teerlink, J.L. Stanford, W.B. Isaacs, J. Xu, K.A. Cooney, E.M. Lange, J. Schleutker, J.D. Carpten, I.J. Powell, O. Cussenot, G. Cancel-Tassin, G.G. Giles, R.J. MacInnis, C. Maier, C.L. Hsieh, F. Wiklund, W.J. Catalona, W.D. Foulkes, D. Mandal, R.A. Eeles, Z. Kote-Jarai, C.D. Bustamante, D.J. Schaid, T. Hastie, E.A. Ostrander, J.E. Bailey-Wilson, P. Radivojac, S.N. Thibodeau, A.S. Whittemore, W. Sieh, REVEL: an ensemble method for predicting the pathogenicity of rare missense variants, Am. J. Hum. Genet. 99 (2016) 877–885.

[10] C.J. Tokheim, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, R. Karchin, Evaluating the evaluation of cancer driver genes, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 14330–14335.

[11] B.J. Raphael, J.R. Dobson, L. Oesper, F. Vandin, Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine, Genome Med. 6 (2014) 5.

[12] E. Porta-Pardo, A. Kamburov, D. Tamborero, T. Pons, D. Grases, A. Valencia, N. Lopez-Bigas, G. Getz, A. Godzik, Comparison of algorithms for the detection of cancer drivers at subgene resolution, Nat. Methods 14 (2017) 782–788.

[13] C.L. Sawyers, A. Hochhaus, E. Feldman, J.M. Goldman, C.B. Miller, O.G. Ottmann, C.A. Schiffer, M. Talpaz, F. Guilhot, M.W. Deininger, T. Fischer, S.G. O'Brien, R.M. Stone, C.B. Gambacorti-Passerini, N.H. Russell, J.J. Reiffers, T.C. Shea, B. Chapuis, S. Coutre, S. Tura, E. Morra, R.A. Larson, A. Saven, C. Peschel, A. Gratwohl, F. Mandelli, M. Ben-Am, I. Gathmann, R. Capdeville, R.L. Paquette, B.J. Druker, Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase II study, Blood 99 (2002) 3530–3539.

[14] A. Hauschild, J.J. Grob, L.V. Demidov, T. Jouary, R. Gutzmer, M. Millward, P. Rutkowski, C.U. Blank, W.H. Miller Jr., E. Kaempgen, S. Martin-Algarra, B. Karaszewska, C. Mauch, V. Chiarion-Sileni, A.M. Martin, S. Swann, P. Haney, B. Mirakhur, M.E. Guckert, V. Goodman, P.B. Chapman, Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial, Lancet 380 (2012) 358–365.

[15] R. Nussinov, C.J. Tsai, H. Jang, Why are some driver mutations rare? Trends Pharmacol. Sci. 40 (2019) 919–929.

[16] H. Takeshima, T. Ushijima, Accumulation of genetic and epigenetic alterations in normal cells and cancer risk, NPJ Precis Oncol 3 (2019) 7.

[17] M.C. King, J.H. Marks, J.B. Mandell, G. New York Breast Cancer Study, Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2, Science 302 (2003) 643–646.

[18] J.R. Pon, M.A. Marra, Driver and passenger mutations in cancer, Annu. Rev. Pathol. 10 (2015) 25–50.

[19] C. Scholl, S. Frohling, Exploiting rare driver mutations for precision cancer medicine, Curr. Opin. Genet. Dev. 54 (2019) 1–6.

[20] M.H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M.C. Wendl, J. Kim, B. Reardon, P.K. Ng, K.J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E.M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortes-Ciriano, D.C. Zhou, W.W. Liang, J.M. Hess, V.D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavilai, J.Y. Ko, E. Khurana, P.J. Park, E.M. Van Allen, H. Liang, M.C.W. Group, N. Cancer Genome Atlas Research, M.S. Lawrence, A. Godzik, N. Lopez-Bigas, J. Stuart, D. Wheeler, G. Getz, K. Chen, A.J. Lazar, G.B. Mills, R. Karchin, L. Ding, Comprehensive characterization of cancer driver genes and mutations, Cell 173 (2018), 371–385 e318.

[21] J.A. Curtin, J. Fridlyand, T. Kageshita, H.N. Patel, K.J. Busam, H. Kutzner, K.H. Cho, S. Aiba, E.B. Brocker, P.E. LeBoit, D. Pinkel, B.C. Bastian, Distinct sets of genetic alterations in melanoma, N. Engl. J. Med. 353 (2005) 2135–2147.

[22] E.R. Fearon, Molecular genetics of colorectal cancer, Annu. Rev. Pathol. 6 (2011) 479–507.

[23] C. Kandoth, M.D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J.F. McMichael, M.A. Wyczalkowski, M.D.M. Leiserson, C.A. Miller, J.S. Welch, M.J. Walter, M.C. Wendl, T.J. Ley, R.K. Wilson, B.J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types, Nature 502 (2013) 333–339.

[24] I.T.P.-C.A.o.W.G. Consortium, Pan-cancer analysis of whole genomes, Nature 578 (2020) 82–93.

[25] O. Harismendy, R.B. Schwab, L. Bao, J. Olson, S. Rozenzhak, S.K. Kotsopoulos, S. Pond, B. Crain, M.S. Chee, K. Messer, D.R. Link, K.A. Frazer, Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing, Genome Biol. 12 (2011) R124.

[26] R. Nussinov, H. Jang, C.J. Tsai, F. Cheng, Review: precision medicine and driver mutations: computational methods, functional assays and conformational principles for interpreting cancer drivers, PLoS Comput. Biol. 15 (2019), e1006658.

[27] A.L. Brown, M. Li, A. Goncearenco, A.R. Panchenko, Finding driver mutations in cancer: elucidating the role of background mutational processes, PLoS Comput. Biol. 15 (2019), e1006981.

[28] T. Dogruluk, Y.H. Tsang, M. Espitia, F. Chen, T. Chen, Z. Chong, V. Appadurai, A. Dogruluk, A.K. Eterovic, P.E. Bonnen, C.J. Creighton, K. Chen, G.B. Mills, K.L. Scott, Identification of variant-specific functions of PIK3CA by rapid phenotyping of rare mutations, Cancer Res. 75 (2015) 5341–5354.

[29] S. Agajanian, O. Odeyemi, N. Bischoff, S. Ratra, G.M. Verkhivker, Machine learning classification and structure-functional analysis of cancer mutations reveal unique

dynamic and network signatures of driver sites in oncogenes and tumor suppressor genes, J. Chem. Inf. Model. 58 (2018) 2131–2150.

[30] D.B. Wetlaufer, Nucleation, rapid folding, and globular intrachain regions in proteins, Proc. Natl. Acad. Sci. U. S. A. 70 (1973) 697–701.

[31] C. Chothia, J. Gough, C. Vogel, S.A. Teichmann, Evolution of the protein repertoire, Science 300 (2003) 1701–1703.

[32] T. Kawashima, S. Kawashima, C. Tanaka, M. Murai, M. Yoneda, N.H. Putnam, D. S. Rokhsar, M. Kanehisa, N. Satoh, H. Wada, Domain shuffling and the evolution of vertebrates, Genome Res. 19 (2009) 1393–1403.

[33] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (1995) 536–540.

[34] L. Lo Conte, B. Ailey, T.J. Hubbard, S.E. Brenner, A.G. Murzin, C. Chothia, SCOP: a structural classification of proteins database, Nucleic Acids Res. 28 (2000) 257–259.

[35] A. Andreeva, D. Howorth, C. Chothia, E. Kulesha, A.G. Murzin, SCOP2 prototype: a new approach to protein structure mining, Nucleic Acids Res. 42 (2014) D310–D314.

[36] M. Blum, H.Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G.A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D.H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, C.A. Orengo, A.P. Pandurangan, C. Rivoire, C.J. A. Sigrist, I. Sillitoe, N. Thanki, P.D. Thomas, S.C.E. Tosatto, C.H. Wu, A. Bateman, R.D. Finn, The InterPro protein families and domains database: 20 years on, Nucleic Acids Res. 49 (2021) D344–D354.

[37] A. Bateman, E. Birney, R. Durbin, S.R. Eddy, K.L. Howe, E.L. Sonnhammer, The Pfam protein families database, Nucleic Acids Res. 28 (2000) 263–266.

[38] I. Sillitoe, N. Bordin, N. Dawson, V.P. Waman, P. Ashford, H.M. Scholes, C.S. M. Pang, L. Woodridge, C. Rauer, N. Sen, M. Abbasian, S. Le Cornu, S.D. Lam, K. Berka, I.H. Varekova, R. Svobodova, J. Lees, C.A. Orengo, CATH: increased structural coverage of functional space, Nucleic Acids Res. 49 (2021) D266–D273.

[39] V. Parthiban, M.M. Gromiha, D. Schomburg, CUPSAT: prediction of protein stability upon point mutations, Nucleic Acids Res. 34 (2006) W239–W242.

[40] J.S. Kaminker, Y. Zhang, A. Waugh, P.M. Haverty, B. Peters, D. Sebisanovic, J. Stinson, W.F. Forrest, J.F. Bazan, S. Seshagiri, Z. Zhang, Distinguishing cancer-associated missense mutations from common polymorphisms, Cancer Res. 67 (2007) 465–473.

[41] H. Carter, S. Chen, L. Isik, S. Tyekucheva, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, R. Karchin, Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations, Cancer Res. 69 (2009) 6660–6667.

[42] M.S. Cline, R. Karchin, Using bioinformatics to predict the functional impact of SNVs, Bioinformatics 27 (2011) 441–448.

[43] Q. Zhong, N. Simonis, Q.R. Li, B. Charloteaux, F. Heuze, N. Klitgord, S. Tam, H. Yu, K. Venkatesan, D. Mou, V. Swearingen, M.A. Yildirim, H. Yan, A. Dricot, D. Szeto, C. Lin, T. Hao, C. Fan, S. Milstein, D. Dupuy, R. Brasseur, D.E. Hill, M.E. Cusick, M. Vidal, Edgetic perturbation models of human inherited disorders, Mol. Syst. Biol. 5 (2009) 321.

[44] L. Shen, Q. Shi, W. Wang, Double agents: genes with both oncogenic and tumor-suppressor functions, Oncogenesis 7 (2018) 25.

[45] N.L. Nehrt, T.A. Peterson, D. Park, M.G. Kann, Domain landscapes of somatic mutations in cancer, BMC Genomics 13 (Suppl. 4) (2012). S9.

[46] B. Reva, Y. Antipin, C. Sander, Predicting the functional impact of protein mutations: application to cancer genomics, Nucleic Acids Res. 39 (2011), e118.

[47] F.S. Krebs, V. Zoete, M. Trottet, T. Pouchon, C. Bovigny, O. Michielin, Swiss-PO: a new tool to analyze the impact of mutations on protein three-dimensional structures for precision oncology, NPJ Precis Oncol 5 (2021) 19.

[48] H.C. Jubb, H.K. Saini, M.L. Verdonk, S.A. Forbes, COSMIC-3D provides structural perspectives on cancer genetics for drug discovery, Nat. Genet. 50 (2018) 1200–1202.

[49] J. Gao, M.T. Chang, H.C. Johnsen, S.P. Gao, B.E. Sylvester, S.O. Sumer, H. Zhang, D.B. Solit, B.S. Taylor, N. Schultz, C. Sander, 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets, Genome Med. 9 (2017) 4.

[50] N. Niknafs, D. Kim, R. Kim, M. Diekhans, M. Ryan, P.D. Stenson, D.N. Cooper, R. Karchin, MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures, Hum. Genet. 132 (2013) 1235–1243.

[51] E. Porta-Pardo, T. Hrabe, A. Godzik, Cancer3D: understanding cancer mutations through protein structures, Nucleic Acids Res. 43 (2015) D968–D973.

[52] S. Hashemi, A. Nowzari Dalini, A. Jalali, A.M. Banaei-Moghaddam, Z. Razaghi-Moghadam, Cancerouspdomains: comprehensive analysis of cancer type-specific recurrent somatic mutations in proteins and domains, BMC Bioinformatics 18 (2017) 370.

[53] G.E. Melloni, S. de Pretis, L. Riva, M. Pelizzola, A. Ceol, J. Costanza, H. Muller, L. Zammataro, LowMACA: exploiting protein family analysis for the identification of rare driver mutations in cancer, BMC Bioinformatics 17 (2016) 80.

[54] T.A. Peterson, A. Adadey, I. Santana-Cruz, Y. Sun, A. Winder, M.G. Kann, DMDM: domain mapping of disease mutations, Bioinformatics 26 (2010) 2458–2459.

[55] T.A. Peterson, I.I.M. Gauran, J. Park, D. Park, M.G. Kann, Oncodomains: a protein domain-centric framework for analyzing rare variants in tumor samples, PLoS Comput. Biol. 13 (2017), e1005428.

[56] T.A. Peterson, N.L. Nehrt, D. Park, M.G. Kann, Incorporating molecular and functional context into the analysis and prioritization of human variants associated with cancer, J. Am. Med. Inform. Assoc. 19 (2012) 275–283.

[57] P. Yue, W.F. Forrest, J.S. Kaminker, S. Lohr, Z. Zhang, G. Cavet, Inferring the functional effects of mutation through clusters of mutations in homologous proteins, Hum. Mutat. 31 (2010) 264–271.

[58] F. Yang, E. Petsalaki, T. Rolland, D.E. Hill, M. Vidal, F.P. Roth, Protein domain-level landscape of cancer-type-specific somatic mutations, PLoS Comput. Biol. 11 (2015), e1004147.

[59] M.L. Miller, E. Reznik, N.P. Gauthier, B.A. Aksoy, A. Korkut, J. Gao, G. Ciriello, N. Schultz, C. Sander, Pan-cancer analysis of mutation hotspots in protein domains, Cell Systems 1 (2015) 197–209.

[60] N.P. Gauthier, E. Reznik, J. Gao, S.O. Sumer, N. Schultz, C. Sander, M.L. Miller, MutationAligner: a resource of recurrent mutation hotspots in protein domains in cancer, Nucleic Acids Res. 44 (2016) D986–D991.

[61] E. Grillo, M. Corsini, C. Ravelli, M. di Somma, L. Zammataro, E. Monti, M. Presta, S. Mitola, A novel variant of VEGFR2 identified by a pan-cancer screening of recurrent somatic mutations in the catalytic domain of tyrosine kinase receptors enhances tumor growth and metastasis, Cancer Lett. 496 (2021) 84–92.

[62] R.A. Toledo, E. Garralda, M. Mitsi, T. Pons, J. Monsech, E. Vega, A. Otero, M. I. Albarran, N. Banos, Y. Duran, V. Bonilla, F. Sarno, M. Camacho-Artacho, T. Sanchez-Perez, S. Perea, R. Alvarez, A. De Martino, D. Lietha, C. Blanco-Aparicio, A. Cubillo, O. Dominguez, J.L. Martinez-Torrecuadrada, M. Hidalgo, Exome sequencing of plasma DNA portrays the mutation landscape of colorectal cancer and discovers mutated VEGFR2 receptors as modulators of antiangiogenic therapies, Clinical Cancer Res. 24 (2018) 3550–3559.

[63] E. Grillo, M. Corsini, C. Ravelli, L. Zammataro, M. Bacci, A. Morandi, E. Monti, M. Presta, S. Mitola, Expression of activated VEGFR2 by R1051Q mutation alters the energy metabolism of Sk-Mel-31 melanoma cells by increasing glutamine dependence, Cancer Lett. 507 (2021) 80–88.

[64] T. Yu, K.P. Choi, E.S. Chen, L. Zhang, Stage-specific protein-domain mutational profile of invasive ductal breast cancer, BMC Med. Genet. 13 (2020) 150.

[65] P. Ashford, C.S.M. Pang, A.A. Moya-Garcia, T. Adeyelu, C.A. Orengo, A CATH domain functional family based approach to identify putative cancer driver genes and driver mutations, Sci. Rep. 9 (2019) 263.

[66] J.T. Lee, L. Li, P.A. Brafford, M. van den Eijnden, M.B. Halloran, K. Sproesser, N. K. Haass, K.S. Smalley, J. Tsai, G. Bollag, M. Herlyn, PLX4032, a potent inhibitor of the B-Raf V600E oncogene, selectively inhibits V600E-positive melanomas, Pigment Cell Melanoma Res 23 (2010) 820–827.

[67] K.T. Flaherty, I. Puzanov, K.B. Kim, A. Ribas, G.A. McArthur, J.A. Sosman, P. J. O'Dwyer, R.J. Lee, J.F. Grippo, K. Nolop, P.B. Chapman, Inhibition of mutated, activated BRAF in metastatic melanoma, N. Engl. J. Med. 363 (2010) 809–819.