ARTICLE   OPEN

Check for updates

The *Cancermuts* software package for the prioritization of missense cancer variants: a case study of AMBRA1 in melanoma

Matteo Tiberti [1,8✉], Luca Di Leo [2,8], Mette Vixø Vistesen[3], Rikke Sofie Kuhre[2], Francesco Cecconi[3,4,5], Daniela De Zio [2,6,9✉] and Elena Papaleo[1,7,9]

Cancer genomics and cancer mutation databases have made an available wealth of information about missense mutations found in cancer patient samples. Contextualizing by means of annotation and predicting the effect of amino acid change help identify which ones are more likely to have a pathogenic impact. Those can be validated by means of experimental approaches that assess the impact of protein mutations on the cellular functions or their tumorigenic potential. Here, we propose the integrative bioinformatic approach *Cancermuts*, implemented as a Python package. *Cancermuts* is able to gather known missense cancer mutations from databases such as cBioPortal and COSMIC, and annotate them with the pathogenicity score REVEL as well as information on their source. It is also able to add annotations about the protein context these mutations are found in, such as post-translational modification sites, structured/unstructured regions, presence of short linear motifs, and more. We applied *Cancermuts* to the intrinsically disordered protein AMBRA1, a key regulator of many cellular processes frequently deregulated in cancer. By these means, we classified mutations of AMBRA1 in melanoma, where AMBRA1 is highly mutated and displays a tumor-suppressive role. Next, based on REVEL score, position along the sequence, and their local context, we applied cellular and molecular approaches to validate the predicted pathogenicity of a subset of mutations in an in vitro melanoma model. By doing so, we have identified two AMBRA1 mutations which show enhanced tumorigenic potential and are worth further investigation, highlighting the usefulness of the tool. *Cancermuts* can be used on any protein targets starting from minimal information, and it is available at https://www.github.com/ELELAB/cancermuts as free software.

INTRODUCTION

Recent advances in cancer genomics have been leading to increased information on cancer mutations. Resources as Genomic Data Commons (GDC) [1] store information from different studies from cancer genomic initiatives, such as The Cancer Genome Atlas [2] and the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) initiative (https://ocg.cancer.gov/programs/target). Databases such as the Catalogue of Somatic Mutations in Cancer (COSMIC) [3] or cBioPortal [4, 5] are also a useful resources to mine cancer mutations. Providing annotations and predictions, which may help discriminate between mutations with or without a pathogenic impact, is still an open challenge [6–9] and a field in need of urgent investigation. This could be assessed by experimental approaches to determine the impact on protein cellular functions or the tumorigenic potential deriving from the alteration. It is also noteworthy that genomic-related changes in coding regions leading to amino acidic substitution(s) can possibly result in alterations of the protein product in terms of stability, key post-translational modifications regulating protein function, or even interactions with other proteins. Bioinformatic tools have been provided to support annotating of some of these properties, even though not in a systematic manner, and they are often based on web servers and leave little space for customizing the analyses [10–14]. Structure-based methods can be applied to assess these different functional layers [15–21] even though they have a limited scope, especially if the target protein includes large intrinsically disordered regions or regions enriched in low-complexity sequence patterns. Furthermore, the context in which any mutation is found is also relevant, as it can be indicative of putative effects of a mutation. For instance, it can be useful knowing whether a certain substitution falls within a binding site for another protein, or whether it is located within a structured region, or whether the substitution can abolish or even introduce a new post-translational modification.

[1]Cancer Structural Biology, Center for Autophagy, Recycling and Disease (CARD), Danish Cancer Society Research Center, Copenhagen, Denmark. [2]Melanoma Research Team, Center for Autophagy, Recycling and Disease (CARD), Danish Cancer Society Research Center, Copenhagen, Denmark. [3]Cell Stress and Survival, Center for Autophagy, Recycling and Disease (CARD), Danish Cancer Society Research Center, Copenhagen, Denmark. [4]Department of Biology, University of Rome Tor Vergata, Rome, Italy. [5]Department of Pediatric Onco-Hematology and Cell and Gene Therapy, IRCCS Bambino Gesù Children's Hospital, Rome, Italy. [6]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [7]Cancer Systems Biology, Health and Technology Department, Section for Bioinformatics, Technical University of Denmark, Lyngby, Denmark. [8]These authors contributed equally: Matteo Tiberti, Luca Di Leo. [9]These authors jointly supervised this work: Daniela De Zio, Elena Papaleo. ✉email: tiberti@cancer.dk; dzio@cancer.dk
Edited by Professor Gerry Melino

To give an easily accessible overview of (i) the distribution of mutations in a gene, (ii) pathogenicity scores, and (iii) annotations along the protein sequence, we have created *Cancermuts*, a Python package that streamlines the collection and annotations of cancer mutations located in the coding region of a gene of interest, e.g. mutations that will impact its protein product. The information is superimposed with different layers to help make informed decisions on which mutations are more likely to be functionally damaging and worth further investigation by either computational or experimental approaches. Cancermuts is approachable for users with minimal Python or programming experience. Nonetheless, this makes it possible to easily integrate it in more complex workflows and grants a high degree of flexibility, customizability, and extendibility.

To validate *Cancermuts* in a cancer study, we focused on the tumor suppressor gene *AMBRA1* (autophagy and beclin 1 regulator 1). Initially discovered to be involved in correct embryogenesis, especially during brain development, in mouse congenital malformations as well as in human neurological disorders [22, 23], AMBRA1 is mostly known for its role in autophagy activation [22, 24]. New cancer-related roles for AMBRA1 have emerged over the years, particularly with regard to cell proliferation and tumorigenic potential [25]. More recently, the role of AMBRA1 as tumor suppressor has been further extended, as by its regulation of cell cycle by Cyclin D1 stabilization (via interaction with the E3 Ubiquitin ligase DDB1-Cullin4) [26–28] and of malignant invasiveness (through focal adhesion kinase FAK1 hyperactivation) [29]. Such a vast range of functions is intertwined with the ability of AMBRA1 to interact with molecular partners [22, 24–28, 30–38] and undergo post-translational modifications (PTMs) [24, 39, 40], and deeply relies on its intrinsically disordered structure.

In this study, we used *Cancermuts* on AMBRA1, allowing to identify putative cancer mutations of interest to be further validated experimentally. The prediction of pathogenic mutations of AMBRA1 and their in vitro validation have been carried out in melanoma, the most aggressive and lethal form of skin cancer, in which AMBRA1 not only bears an anti-tumorigenic function [29], but also displays high mutation rate [26].

RESULTS
Design and implementation
Cancermuts is designed as a Python package with an easy-to-use programming interface (API) (Fig. 1). It is suited to researchers with basic programming Python skills and can be used, for instance, in popular interactive Python interfaces, such as Jupyter notebooks, as well as in standalone Python scripts, or integrated in more complex workflows. In fact, the information obtained from *Cancermuts* can be represented as a Pandas data frame, a commonly used data format that can be easily further processed for data exploration or integration with other sources. The *Cancermuts* API also allows a good degree of flexibility, allowing the researcher to customize several aspects on which information is collected and to build their own annotation strategy.

Cancermuts only requires basic information about the gene of interest, namely either its IDs or its protein product IDs, such as Uniprot accession ID [41]. Using the *Cancermuts* API, the user is expected to download the protein sequence first by providing relevant database IDs (Fig. 1). This can then be annotated with protein missense mutations from cancer mutation or genomics databases, as well as with further annotations regarding the protein sequence itself and the identified mutations (see below for details). Both mutations, e.g. from patient-cohort studies, and annotations can also be provided from custom user-designed input files, allowing integration in the annotation pipeline. The tool has been designed to annotate somatic mutations and especially focuses on single nucleotide variants. Once the data

collection has been performed, *Cancermuts* provides the researcher with a textual and graphical representation of the mutations to explore the data and help with their interpretation. All the obtained data can be converted to a simple Pandas data frame, which can then be manipulated as desired, including saving it as a table (CSV) file. *Cancermuts* also includes facilities to represent the annotation as a publication-ready stem plot which can be thoroughly customized.

Cancermuts interacts with different freely available resources for sequence-based annotations as detailed below. The package is designed to be modular and easily extendable, should other annotations be of interest in the future.

The current release interacts with the cancer databases COSMIC [3] and cBioPortal [4, 5] to retrieve cancer-associated mutations, allowing local or on-the-fly access, respectively, along with data filtering starting from minimal information about the gene of interest. It is also possible to filter for cancer type or study. Some of the metadata from these databases are kept as annotations.

In addition, *Cancermuts* retrieves a score (ranging between 0 and 1) for pathogenicity of the mutations based on 13 individual predictors that have been combined using a random forest approach within the REVEL framework [42]. The user can deduce the threshold value to associate with pathogenic mutations based on specific case studies and benchmarking. We recommend applying a cut-off of 0.4 in case additional information are lacking, i.e. the one that guarantees a good compromise between specificity (0.85) and sensitivity (0.81) [42].

The tool is also able to query gnomAD [43], a collection of harmonized whole-genome and -exome sequencing data. gnomAD works as a proxy of the healthy population for allele frequency. This annotation can be used, for example, to discard some of the mutations from further analyses. Indeed, if a mutation occurs with high frequency in non-tumoral samples, it is unlikely to have a strong pathogenic impact.

Cancermuts allows to store annotations for functional short-linear motifs (SLiMs) that might be related to protein regulation or interaction. This is done interacting with the ELM database [44] or providing input annotations from other sources. Information on putative PTMs are provided by querying PhosphoSitePlus [45] and can additionally be provided by the user in case additional annotations not covered in the database (but experimentally proven) are available.

Cancermuts also allows to annotate the propensity to disorder or structure using MobiDB. Additional custom annotations regarding structure propensity can be provided by the user through a specific formatted CSV file (see GitHub repository and user guide).

Cancermuts has been designed to be applicable to any protein product and does not require structural information. It is especially interesting for intrinsically disordered proteins or domains, along with low complexity repeats for which structure-based methods currently available to predict the effect of mutations are not easily transferable or challenging to apply. Structural annotations can be facultatively added, whether available.

Cancermuts can be used on any protein target and it is available at https://www.github.com/ELELAB/cancermuts as free software accompanied by a tutorial on GitBook that details its usage on another protein target (i.e., LC3B).

Case study: AMBRA1 mutations in melanoma
AMBRA1 is a large, mostly disordered protein, with a canonical UniProt sequence of 1298 residues. The intrinsically disordered nature of the protein, along with its high plasticity, probably due to its several protein-protein interactions and post-translational modifications (PTMs) [22–40], make of AMBRA1 a good candidate to link together different intracellular processes. Notably, many types of cancer, including malignant melanoma, where AMBRA1 has been shown to play an anti-tumorigenic role [29], show
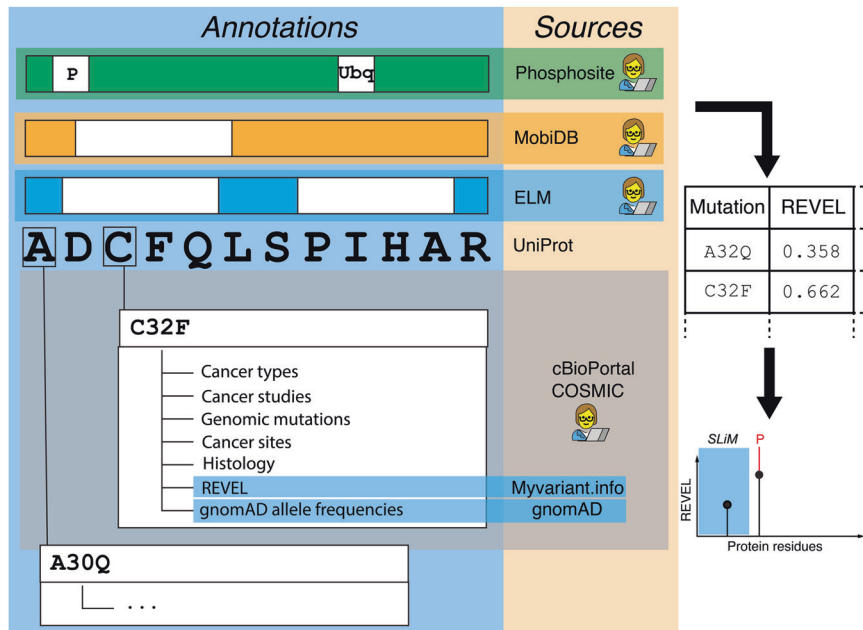
Fig. 1 Schematic representation of the Cancermuts workflow. The figure shows on the left the different layers of evidence that Cancermuts supports. The Uniprot main isoform sequence is the basis on which either per-position or per-sequence annotations are performed (post-translational modifications, structure definition, linear motifs) that can be provided by one of the sources and by manual user annotation. The sequence can also be annotated by downloading cancer-related mutations and relative metadata, including REVEL scores and gnomAD allele frequencies. Mutations can be supplied by manual annotation as well. Once the information has been collected, it can be summarized as a table (right) and as a plot (bottom right).

genetic alterations in *AMBRA1* [26–28]. Indeed, AMBRA1—when compared to other cancer studies, shows the highest mutation rate in skin cancers, such as melanoma [26].

Due to its structure, propensity to interact with other proteins, and cancer-related functions [26–29], we sought to apply Cancermuts to *AMBRA1* in order to predict the pathogenicity of its mutations in melanoma.

Cancermuts for AMBRA1: in silico analysis

We have collected all available melanoma-associated mutations for AMBRA1 from COSMIC and mutations associated with melanoma studies from cBioPortal on 8 April 2020. We have annotated this information with all the available annotations in *Cancermuts* as well as integrated them with manual annotations. These include information about SLiM binding sites and PTMs known in literature but unavailable on the databases on the date the pipeline was run, as well as more about predicted structural regions of AMBRA1 (see GitHub repository). By using a model based on AlphaFold2, we have predicted residues ~7–203 and 857–1040 (381 residues: ~29% of total sequence length) of AMBRA1 to be structured regions, including a region with a β-propeller fold (Fig. 2A). We have saved all the collected information in a CSV table (see Supplemental Table S1) and provided a graphical support (Fig. 2B).

Overall, our analysis identified a total of 73 melanoma-associated non-neutral mutations (Supplemental Table S2, Fig. 2B), 70 of which derived from single-nucleotide substitutions and 3 from multiple nucleotide substitutions (P589F, S605F, P1253S). As the REVEL score is only available for single substitution, these could not be assigned any score. Of such mutations, 40% (28) displayed a REVEL score below significant threshold (Supplemental Table S2, Fig. 2B). Out of the identified protein mutations, the genomic alterations associated to 54 (~74%) were found compatible with UV-induced DNA damage (Supplemental Table S3). Overall, about 39% of the identified mutations were found to be in putative structured regions, displaying no general

preference for such regions to accumulate mutations. Nonetheless, out of the 28 mutations predicted as damaging for REVEL, the majority (15) was found within the predicted structure regions, whereas only 13 were found within the disordered parts of the protein, which covers nearly 60% of the sequence. Therefore, at least for this specific case, a damaging mutation is more likely to be found in a predicted structured region. Pro and Ser were by far the most mutated residue types (17 for each, respectively), followed by Arg (8), Gly (8), and Leu (7). Not surprisingly, the most frequent mutation in the dataset was Ser to Phe (10 occurrences) and Pro to Ser (9), followed by Pro to Leu (5) and Leu to Phe (5), with all the other substitutions being far less frequent (two occurrences or less). The mutation frequencies downloaded by gnomAD did not allow us to discard any of the identified mutations in this case.

We have annotated a total of 26 phosphorylation sites, 6 ubiquitylation sites, and 7 methylation sites. Most of the PTMs sites are localized in unstructured regions of the protein, where they could be more accessible to kinases or other proteins responsible for their modification. Phosphorylations tend to be clustered in small groups, for instance in stretches 387–394, 628–639, 797–814, 1027–1043, 1192–1205, which may be important regulatory regions of the protein [24, 25, 31, 33, 34, 40]. Ubiquitylation sites are clustered in the first 50 residues of AMBRA1, while methylation sites are found mostly in the 730–824 region.

ELM identified several potential SLiMs to which interactors could bind close to mutation sites. It should be noted that SLiMs are defined in the context of disordered regions, while the SLiMs identified in *Cancermuts* are not filtered according to *Cancermuts*' own definition of structured or unstructured regions, as such information could still be useful, depending on how trustable the disorder prediction or definition is, as well as considering order-disorder transitions in the protein structure. Interestingly, ELM identified different TRAF6 ubiquitin ligase binding sites, the role of which in AMBRA1-mediated control of autophagy has already been described [24]. Other relevant binding sites include those for cyclins.

A

B
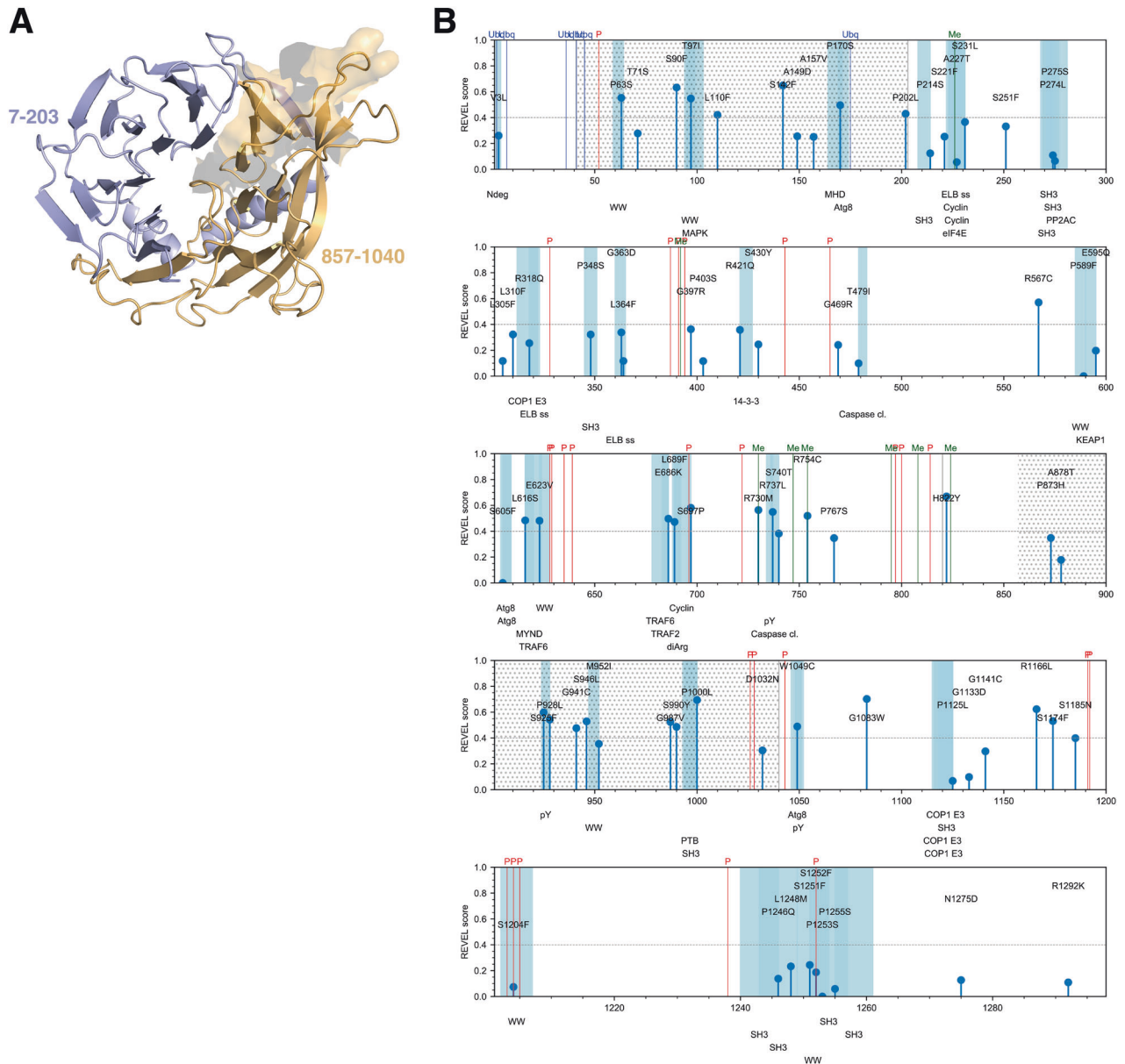
Fig. 2 Predicted structured regions of AMBRA1 and identified cancer mutations. A predicted structured regions of AMBRA1 including the β-propeller domain. The N-terminal and C-terminal regions of the model are shown in blue (7–203) and orange (857–1040), respectively. In the N-terminal region, we included in the model the helical structures (residues 7–41) that are upstream of the propeller domain since they were predicted with high confidence by AlphaFold2. The part of this structured region against which the AMBRA1 antibody for the C-terminal region we have used has been raised is also highlighted by showing its molecular surface (top right). **B** Melanoma-related mutations and corresponding annotations as collected by Cancermuts. Plots follow the main Uniprot sequence numbering; each mutation is annotated as a stem the height of which is proportional to the identified REVEL score (with 0 for those that could not be annotated). Blue shades and corresponding bottom labels refer to linear motifs as annotated by ELM. Post-translational modifications are shown as colored vertical lines. A gray dotted pattern was used to represent the predicted structured β-propeller domain of AMBRA1. Predicted SLIMs that do not overlap with mutations have been hidden from this plot for ease of visualization.
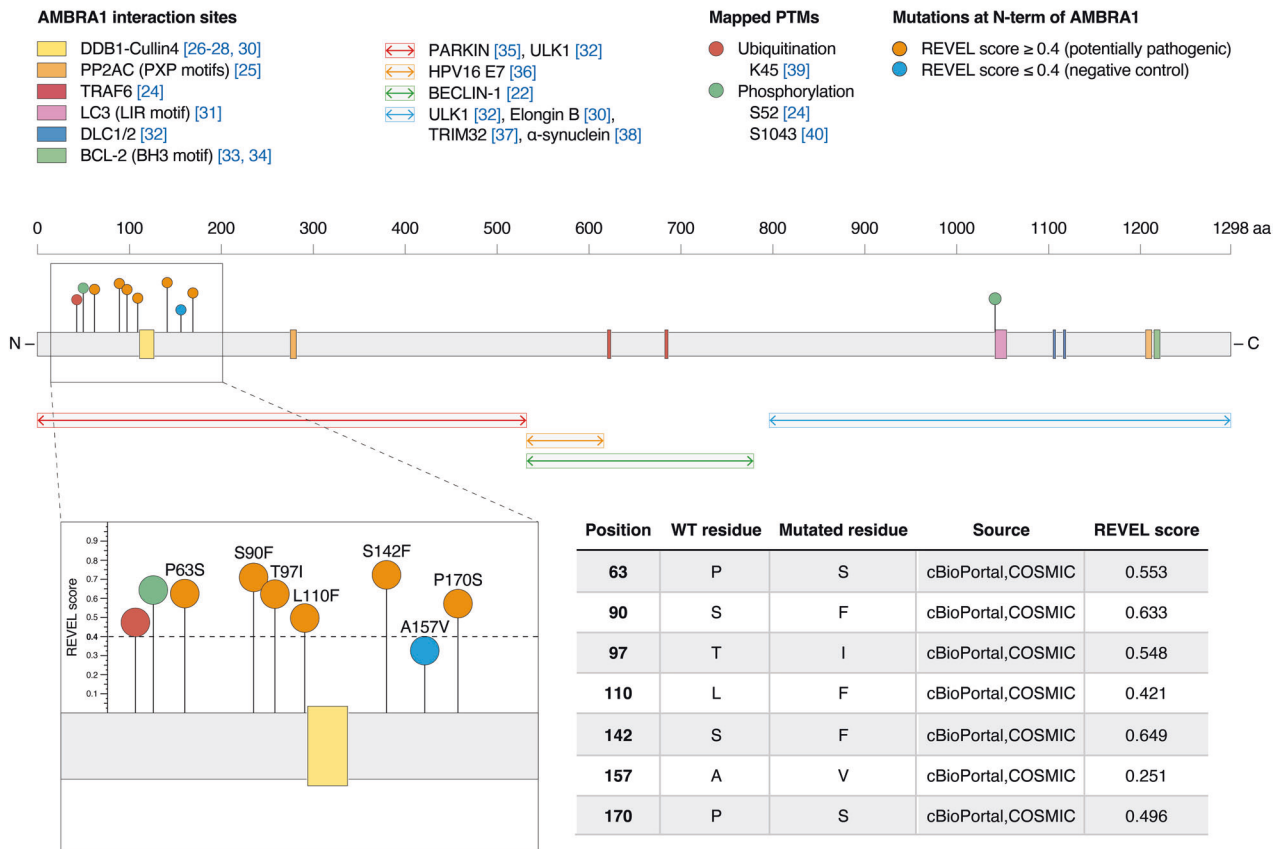
Cancermuts for AMBRA1: in vitro validation

Among the identified mutations, several of those predicted as pathogenic are included both in the N- and C-terminal predicted β-propeller regions (Fig. 2B). Based on the recent findings indicating that the N-terminus of AMBRA1 is involved in stabilization of AMBRA1 itself [30] and of Cyclin D1 [26–28], a result that we confirmed in BRAF^{V600E}-mutated A375 melanoma cells silenced for *AMBRA1* by small interference RNA (siRNA; siAMBRA1 #1 and #2) (Supplemental Fig. 1), we characterized the in vitro effects of AMBRA1 mutations (REVEL score ≥0.4) within those mapping at the N-terminus of the protein (Fig. 3A). The list
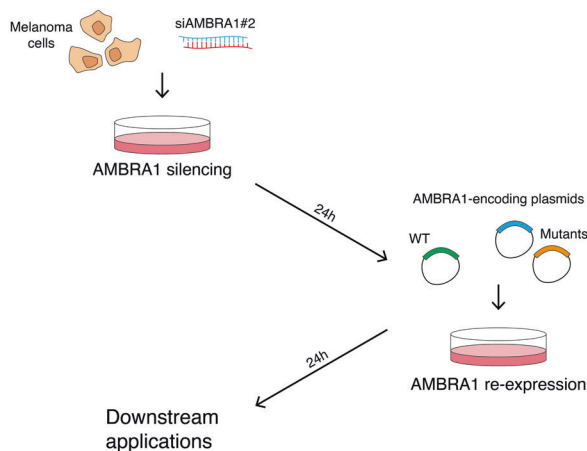
of these mutations, as well as interaction sites [24–28, 30–34] and PTMs [22, 30, 32, 35–40] of AMBRA1 are shown in Fig. 3A. Our analyses also include the A157V mutation which, bearing a REVEL score ≤ 0.4, and an amino acid substitution with a residue of similar type and steric incumbrance (A to V), is not predicted to be pathogenic. Re-expression of WT AMBRA1 has been used as a reference. Our experimental settings consist of transfecting melanoma cells with a siRNA targeting the 5′-UTR region of *AMBRA1* (siAMBRA1#2) prior to mutant re-expression (Fig. 3B). Western blot (WB) analyses ruled out any possible effects of the mutated constructs on either the autophagy or apoptotic
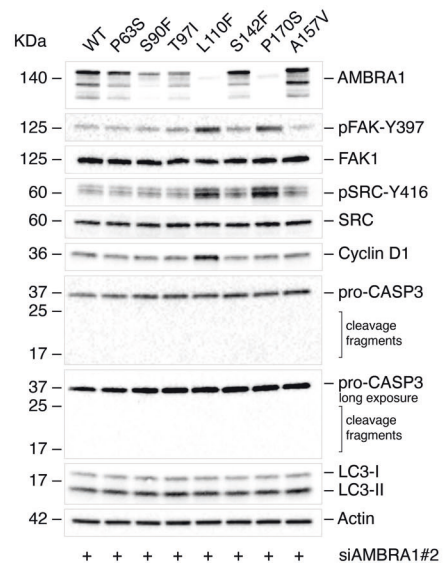
A

AMBRA1 interaction sites

- DDB1-Cullin4 [26-28, 30]
- PP2AC (PXP motifs) [25]
- TRAF6 [24]
- LC3 (LIR motif) [31]
- DLC1/2 [32]
- BCL-2 (BH3 motif) [33, 34]

- ↔ PARKIN [35], ULK1 [32]
- ↔ HPV16 E7 [36]
- ↔ BECLIN-1 [22]
- ↔ ULK1 [32], Elongin B [30], TRIM32 [37], α-synuclein [38]

Mapped PTMs

- Ubiquitination K45 [39]
- Phosphorylation S52 [24] S1043 [40]

Mutations at N-term of AMBRA1

- REVEL score ≥ 0.4 (potentially pathogenic)
- REVEL score ≤ 0.4 (negative control)

Position	WT residue	Mutated residue	Source	REVEL score
63	P	S	cBioPortal,COSMIC	0.553
90	S	F	cBioPortal,COSMIC	0.633
97	T	I	cBioPortal,COSMIC	0.548
110	L	F	cBioPortal,COSMIC	0.421
142	S	F	cBioPortal,COSMIC	0.649
157	A	V	cBioPortal,COSMIC	0.251
170	P	S	cBioPortal,COSMIC	0.496

B

Melanoma cells + siAMBRA1#2 → AMBRA1 silencing → (24h) AMBRA1-encoding plasmids WT / Mutants → AMBRA1 re-expression → (24h) Downstream applications

C

Fig. 3 Analysis of the effect(s) of the point mutations of AMBRA1 on its functions. A Schematic representation of AMBRA1 including interaction sites and PTMs identified experimentally. AMBRA1 mutations are also mapped and specified in the table to the right. **B** Schematic representation of the transfection strategy. **C** WB analysis of pFAK-Y397, FAK1, pSRC-Y416, SRC, Cyclin D1, pro-CASP3 (including cleaved fragments), and LC3 (-I and -II) in A375 melanoma cells re-expressing the P63S, S90F, T97I, L110F, S142F, and P170S AMBRA1 mutants. Re-expression of WT and A157V AMBRA1 was used as a reference and negative control, respectively. AMBRA1 and Actin were detected as transfection and loading control, respectively. Images are representative of $n = 3$ independent experiments.

functions of AMBRA1, as respectively stated by lipidation of LC3 (LC3-II) (Fig. 3C; Supplemental Fig. 2A), a bona fide marker for autophagy, and cleavage of the apoptotic marker CASP-3 (Fig. 3C). Instead, increased protein levels of Cyclin D1 were observed solely upon re-expression of the L110F mutant (Fig. 3C; Supplemental Fig. 2B). In addition, re-expression of L110F, and of P170S as well, resulted in hyperphosphorylation of FAK1 at Y397 (pFAK-Y397) and of SRC (another component of the FAK1 signaling) at Y416 (pSRC-Y416), suggesting an active FAK1 signaling in both conditions (Fig. 3C; Supplemental Fig. 2C and D). Interestingly, re-expression of both L110F (close to the DDB1-Cullin4 domain) and P170S (close to a predicted ubiquitination site) mutants resulted in poor detection of AMBRA1 constructs at protein level (Fig. 3C; Supplemental Fig. 2E). On the other hand, no differences were depicted at mRNA level by RT-qPCR with respect to WT-expressing cells (Fig. 4A), hence suggesting possible effects on protein stability. Previously, AMBRA1/DDB1-Cullin4 interaction was shown to promote AMBRA1 stability by proteasome degradation [30]. To assess whether reduced AMBRA1 protein levels could result from protein degradation by either the proteasome or lysosome pathway, L110F- and P170S-expressing A375 cells were treated with proteasome (MG-132) (Fig. 4B) and lysosome (chloroquine, CQ) (Fig. 4C) inhibitors, respectively. However, in neither condition a rescue of protein levels was observed. The presence of protein aggregates was also assessed in insoluble fractions of mutant-expressing cells, however unsuccessfully (Fig. 4D). Interestingly, when other antibodies raised against AMBRA1 were employed, protein levels of L110F and P170S could be successfully detected with an antibody raised against the N-terminus of AMBRA1 (Fig. 4E). Moreover, when myc-tagged AMBRA1 constructs were re-expressed instead (Fig. 4F), and protein levels detected using either an anti-myc or the panel of anti-AMBRA1 antibodies, the expression of the L110F and P170S mutants could be detected in all instances and was comparable to WT-expressing cells. Hence this suggests a failure of the two antibodies raised against the C-terminus of AMBRA1 shown in Fig. 4E, rather than effects of the mutants on protein stability. Functionally, the levels of Cyclin D1, pFAK-Y397, and pSRC-Y416 upon re-expression of myc-tagged mutants were also consistent with those observed in non-myc-tagged expressing cells (Fig. 4G; Supplemental Fig. 2F–H).

The increased levels of Cyclin D1 and the hyperactivation of FAK1 signaling upon *Ambra1* depletion have been previously correlated to boosted proliferative rate and invasiveness of melanoma, respectively [29]. Although the increased Cyclin D1 levels observed upon L110F expression (Fig. 3C; Supplemental Fig.
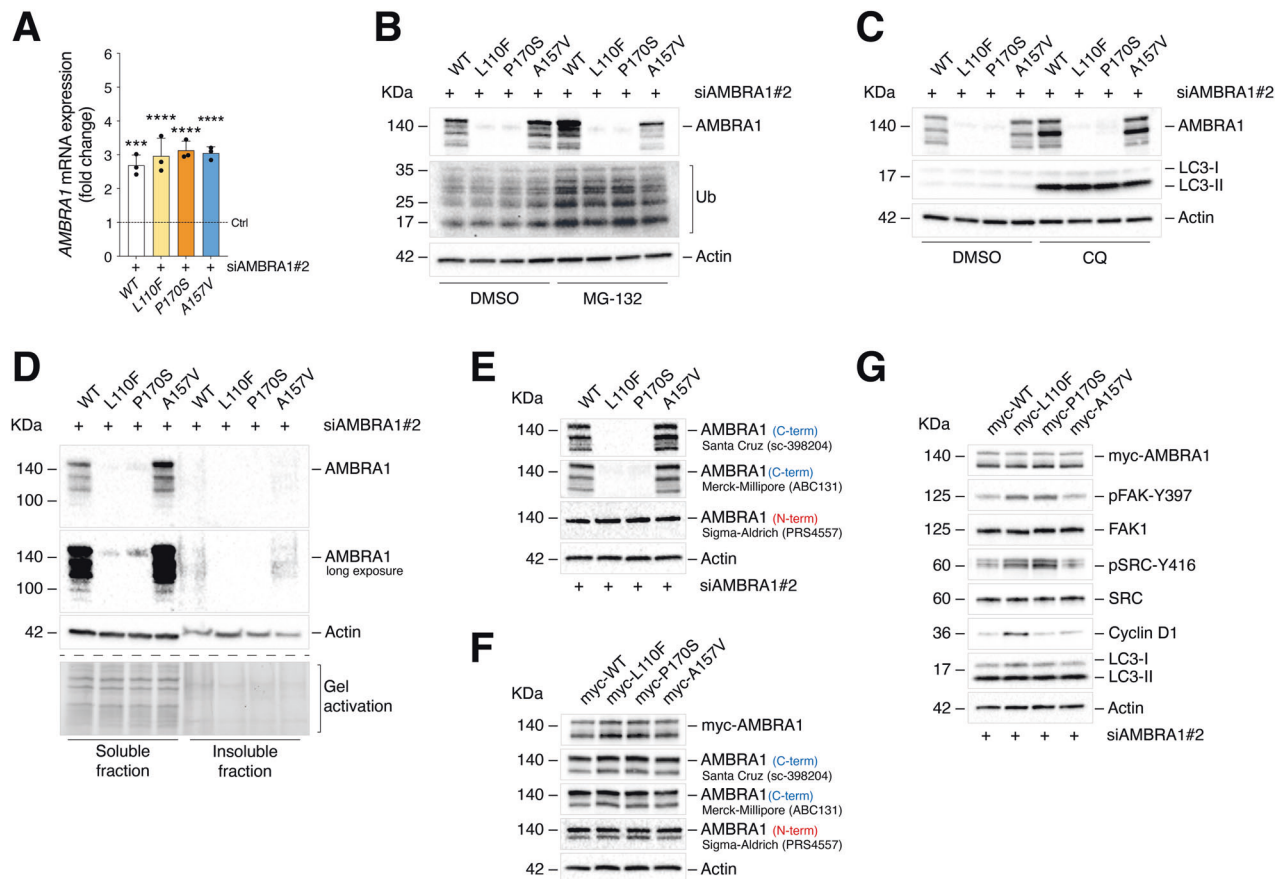
Fig. 4 Analysis of the role of L110F and P170S mutants on AMBRA1 stability. A RT-qPCR analyses of *AMBRA1* upon WT, L110F, P170S, and A157V re-expression. Data were normalized on *L34* and expressed as fold change vs non-transfected cells (Ctrl, indicated by a dashed line) ± SEM ($n = 3$; ***$p = 0.0002$; ****$p < 0.0001$; one-way ANOVA). **B** 24 h after transfection of the constructs, A375 cells were treated with MG-132 (10 µM) or **C** CQ (40 µM) for 4 h and WB analyses for AMBRA1 and Actin performed. Ubiquitylated proteins (Ub) and LC3-II accumulation were detected as positive controls for the treatments. Images are representative of n = 4 independent experiments. **D** WB analyses of soluble and insoluble fractions upon mutant re-expression. AMBRA1 and Actin were detected ($n = 3$). A representative gel activation is also shown. **E** WB analyses of AMBRA1 upon WT, L110F, P170S, and A157V re-expression in A375 and **F** upon re-expression of AMBRA1-myc-tagged constructs using a panel of anti-AMBRA1 antibodies. In **F**, AMBRA1 has been revealed also using an anti-myc antibody. **G** A375 cells were re-expressed with AMBRA1-myc-tagged WT, L110F, P170S and A157V constructs and pFAK-Y397, FAK1, pSRC-Y416, SRC, Cyclin D1, and LC3 (-I and -II) detected by WB. AMBRA1 was detected using an anti-myc antibody. In **E**–**G** Actin was revealed as loading control and images are representative of $n = 3$ independent experiments.
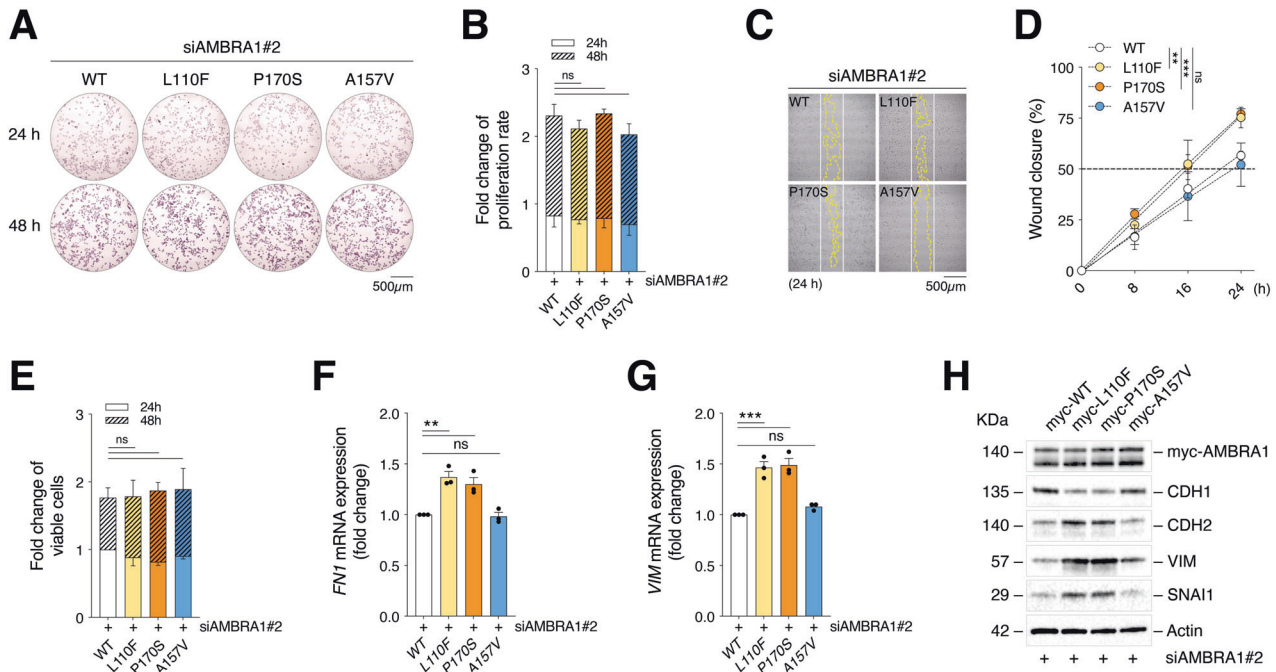
Fig. 5 Analysis of the effect of L110F and P170S mutants on melanoma invasiveness. A Representative Crystal Violet staining of WT-, L110F-, P170S- and A157V-expressing A375 cells after 24 and 48 h ($n = 3$). **B** Quantification of the staining shown in (**A**). Data are shown as fold change ± SD with respect to control sample (WT at 24 h) ($n = 3$; ns=not significant; two-way ANOVA). **C** Representative images ($n = 3$) of wound healing assay in mutant-expressing A375 cells at 24 h. White and yellow lines outline wound edge at T_0 and at the time indicated. **D** Quantification of wound closure is shown as percentage ± SD vs T_0 at the times indicated ($n = 3$; ns=not significant; **$p = 0.0018$; ***$p = 0.0004$; two-way ANOVA). **E** Cell viability of WT-, L110F-, P170S- and A157V-expressing A375 cells after 24 and 48 h is expressed as fold change ± SD with respect to control sample (WT at 24 h) ($n = 3$; ns=not significant; two-way ANOVA). **F** RT-qPCR analyses of EMT markers *FN1* ($n = 3$; ns=not significant; **$p = 0.0016$ L110F vs WT; **$p = 0.006$ P170S vs WT; one-way ANOVA) and **G** *VIM* ($n = 3$; ns not significant; ***$p = 0.0003$ L110F vs WT; ***$p = 0.0002$ P170S vs WT; one-way ANOVA) upon WT, L110F, P170S and A157V re-expression. Data are expressed as fold change vs WT ± SEM. **H** A375 cells were re-expressed with AMBRA1-myc-tagged WT, L110F, P170S and A157V constructs and CDH1, CDH2, VIM and SNAI1 detected by WB. AMBRA1 was detected using an anti-myc antibody as transfection control whereas Actin as loading control. Images are representative of $n = 3$ independent experiments.

2B) did not correlate with changes in proliferation rate of A375 cells (Fig. 5A, B), the hyperphosphorylation of FAK1 and SRC (Fig. 3C; Supplemental Fig. 2C and D) associated with a higher invasive capacity of A375 upon both L110F and P170S re-expression (Fig. 5C, D). No effects were observed upon re-expression of the negative mutant A157V. Such an effect was also unrelated to possible changes in cell viability (Fig. 5E). Consistently with previous data showing that loss of *Ambra1* promotes an epithelial-to-mesenchymal (EMT)-like phenotype in melanoma [29], re-expression of the L110F and P170S also improved expression of the mesenchymal markers *Fibronectin* (*FN1*) (Fig. 5F) and *Vimentin* (*VIM*) (Fig. 5G) at mRNA and of CDH2, VIM and SNAI1 at protein (Fig. 5H; Supplemental Fig. 2I-K) level, whereas a reduced protein expression was observed for the epithelial marker CDH1 (Fig. 5H; Supplemental Fig. 2L).

Comprehensively, our results indicate that, although differently, the expression of the AMBRA1 L110F and P170S mutants, predicted to be pathogenic, accelerates the wound closure capacity of melanoma cells and activates the EMT process and the FAK1 oncogenic signaling pathway.

Prediction of changes in folding free energy upon mutations
We have used an in silico mutational scan based on MutateX [46] and the FoldX energy function [47] to investigate whether AMBRA1 mutations validated for this study were likely to affect the stability of the β-propeller domain (Fig. 6). Our results show that out of the 7 experimentally validated mutations, 4 were mutational hotspots. These are residues 63, 110, 142 and 170 for which the substitution to most residue types was found destabilizing (ΔΔG > 1.2 kcal/mol) (Fig. 6). On the contrary, the

scan shows that any mutation at residue 90 was predicted not to affect stability, while residues 97 and 157 had a less extreme behavior, with only some substitutions having a negative effect (Fig. 6). Unsurprisingly, the experimentally tested mutations at the hotspot sites (P63S, L110F, S142F, P170S) were found to be destabilizing as well (Fig. 6). Mutations S90F and A157V had a neutral effect on stability, while T97I was classified as stabilizing (ΔΔG = −2.34 kcal/mol) (Fig. 6).

DISCUSSION
In this work we have presented *Cancermuts*, a Python package for the discovery, annotation and prioritization of cancer-related mutations. Our software can interrogate cancer genomics and mutation databases such as COSMIC and cBioPortal to retrieve cancer-associated mutations, both in pan-cancer or specific cancer types and studies. It also annotates both protein sequences and identified mutations to (i) give a context in terms of functional or structural features surrounding mutation sites and (ii) assess their potential to interfere with such features. Annotating mutations with pathogenicity scores such as REVEL and gnomAD allele frequencies also helps inform on their potential for pathogenicity more in general.

Cancermuts has been designed as a Python package to ensure maximum flexibility, expandability and modularity. It is straightforward to use for researchers with basic Python skills and can be either used independently or incorporated in more complex workflows, e.g., after reducing the information it collects to a data frame.

In this contribution we have tested our approach on the protein AMBRA1, focusing on cancer mutations from melanoma. Indeed,
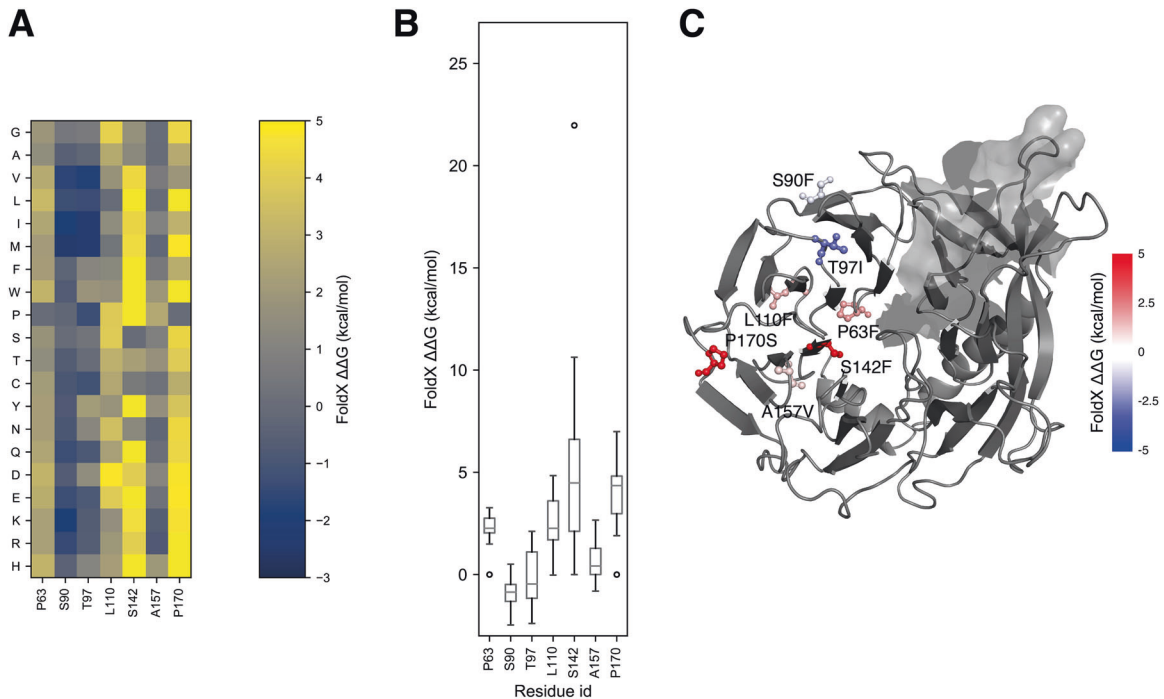
Fig. 6 Predicted changes of folding free energy upon mutation for the mutations experimentally tested. In these plots, positive values represent mutations predicted to be destabilizing the protein structure, while negative values represent mutations predicted to be stabilizing. **A** Heatmap with predicted changes of free energy for the whole mutational scans at the sites we tested experimentally. Values in the plot have been limited within the −3 to 5 kcal/mol range to avoid using a less effective color scale due to outliers. **B** Box plots representing the distribution of the same values plotted in (**A**). **C** Residues at mutation sites corresponding to mutations that have been experimentally tested are shown on the predicted protein structure as sticks, colored proportionally to the predicted free energy change of the respective mutations, as per the colorbar. Values here have been limited in the −5 to 5 kcal/mol for the same reasons as in (**B**) to obtain a symmetrical colormap around 0.

melanoma is one of the cancer types in which AMBRA1 displays a crucial anti-tumorigenic role [29] and a mutational rate among the highest [26]. AMBRA1 is largely an intrinsically disordered protein (IDP), and its "unstructure" suggests that it can adapt to diverse situations and possibly coordinate different intracellular processes mainly by regulating protein-protein interactions [23]. The flexibility of its long-disordered regions could play an important role in modulating the conformational changes needed to provide interaction surfaces that are complementary to different biological partners. Our tool identified several AMBRA1 mutations of potential interest in melanoma and allowed us to contextualize them in terms of localization in a predicted structured region, surrounding post-translational modifications, embedding in short linear motifs, and to annotate them for pathogenicity scores. Based on its importance in AMBRA1 itself [30] and Cyclin D1 [26–28] stability, we then focused our attention on the N-terminal region of the protein and assessed the effect of the most interesting mutations, given their context and annotations. Mutations have been selected by means of the pathogenicity score REVEL using a threshold of 0.4, which corresponds to a good balance between specificity and sensitivity (sensitivity 0.81 and specificity 0.85) [42] and represents a good compromise. Nonetheless, further fine-tuning of the cut-off might be necessary to suit different cases, also depending on the amount of available resources for further validation. Of the tested mutations, none had a clear effect on the AMBRA1-related autophagic or apoptotic pathways. However, we cannot rule out that such mutations might have effects we did not test for, or that such mutations might be detrimental in other conditions or cell types, or in conjunction with others. In this sense, having a wider range of readouts could help understand whether these false positive mutations can be important. This also highlights a potential downside of using

pathogenicity predictors that are not tailored towards a specific disease or tissue. It has been shown that the performance of variant prioritization predictions varies with diseases phenotype [48], and machine learning models trained on more specific datasets could incorporate more of the cellular context of the identified variants or diseases, improving performance [48]. It should also be noted that all the N-terminal tested mutations fall in a generally very well conserved region of AMBRA1, as demonstrated by our protein sequence alignment among sequences of AMBRA1 orthologs from human, chimpanzee, mouse, rat, bovine, *Xenopus* and zebrafish (see Methods and Supplemental Table S4). As 8 out of 18 of the pathogenicity scores integrated in REVEL are based on conservation, this is probably a contributing factor to the score that REVEL assigns to the mutations in this region. Interestingly, although differently, mutations of the conserved L110 (L→F) and P170 (P→S) residues were found implicated in functions of AMBRA1 recently reported to be relevant in terms of tumor growth and progression [26–29]. Among these, the expression of the L110F mutant (which maps next to the DDB1-Cullin4 interaction domain of AMBRA1), correlated with increased protein levels of Cyclin D1. Despite no difference in terms of proliferation was assessed (a counter-intuitive outcome that may be explained by the high proliferative rate of A375 cells), the high Cyclin D1 levels detected in these circumstances may implicate an impaired ability of AMBRA1 to control Cyclin D1 stability. Moreover, both mutations increased the phosphorylation status of components of FAK1 signaling, namely of FAK1 itself (pFAK1-Y397) and SRC (pSRC-Y416). Under the same conditions, cells expressing our mutants displayed higher cell invasiveness, hence suggesting a potential pathogenic effect for either mutation. RTqPCR analyses of cyclin D1 (*CCND1*) upon mutant re-expression, as well as protein expression analyses

of Cyclin D1 in *Ambra1*-null tumors upon FAK1 inhibitor ruled out any correlation between FAK1 signaling activation and Cyclin D1 expression [49] (Supplemental Fig. 3). Structure-based mutational scans suggest that both mutations are likely to destabilize the protein structure. Indeed, both positions were found to be extremely sensitive to mutation in a way that any amino acid change is likely to destabilize the protein at these positions. Despite the protein levels of the L110F and P170S mutants were not affected when screened using myc-tagged AMBRA1 mutants or an anti-N-terminus-AMBRA1 antibody, this does not rule out that the protein may undergo PTMs or that the structure itself might be affected –as suggested by the anti-C-terminus-AMBRA1 antibody failure, which was raised against residues 999–1298 of the AMBRA1 sequence. This includes part of the predicted β-propeller domain that bridges to the N-terminal region by means of a β-sheet. Residues 110 and 170 are not directly in contact with this region (Fig. 6), meaning it is unlikely their mutation would have a direct effect; nonetheless, they could still elicit a long-range effect by disrupting the local propeller structure and interfering with propeller assembly.

Even though these mutations feature the lowest REVEL score among those classified as pathogenic, they were found to bear the largest effects among those we tested. We speculated this might be due to their potential of inducing conformational changes or destabilization of the AMBRA1 protein structure. In this case, therefore, as REVEL does not include predictions of changes in protein structural stability explicitly, additional annotations that rely directly on structural information could complement and add another compelling layer of evidence. In this sense, tools able to perform high-throughput mutational scans (e.g. MutateX, which uses FoldX) aiming at predicting the impact of mutations on protein structure could be integrated in the *Cancermuts* package, for instance by including ready-made mutational scans in the annotation pipeline, which can be provided through the structure-based framework introduced in the work by Fas et al. [15].

Cancermuts was created with a modular design philosophy. This makes it possible to add additional layers of evidence by including support for them in the code, taking advantage of the pre-existing package structure. This will be useful to add new predictors or other data as they become available or to tailor its use to specific cases or datasets. For example, predicted structures from the AlphaFold protein structure models collection [50] could be used to integrate an additional layer of information about the structure and differentiate between predicted disordered and ordered regions. Other attractive layers of evidence also include predictors for the effect of mutation based on sequence evolution, such as the recently released EVE model [51] which relies on generative models of evolutionary data, GEMME [52], DeepSequence [53] or EVmutation [54]. The fact that *Cancermuts* also allows user-curated input yet adds another layer of flexibility, allowing to add information without need to write any code.

MATERIALS AND METHODS
Cell lines and treatments
The human melanoma cell line A375 (RRID: CVCL_0132) was cultivated in GlutaMAX™-additioned Dulbecco's Modified Eagle Medium (DMEM) (ThermoFisher Scientific; cat# 31966-021) supplemented with 10% FBS (ThermoFisher Scientific; cat# 10270-106) and 100 U/ml P/S (Thermo-Fisher Scientific; cat# 15140122). Cells were cultured in an atmosphere of 5% CO_2 in air at 37 °C and passaged no more than 15 times. Cells were used within a few months of resuscitation and routinely tested for *Mycoplasma* during sub-cultivation by PCR-based methods (eurofins Genomics, DE) and only used if negative. During the experiments, cells were plated at a density of 1×10^5 cells/ml, unless otherwise indicated. Chloroquine (CQ, Sigma-Aldrich; cat# C6628) and MG-132 (Sigma-Aldrich; cat# M7449) were dissolved in DMSO and used at final concentrations of 40 and 10 μM, respectively, for 4 h while DMSO was used in control cells.

In vivo analyses
Samples for in vivo analyses have been collected as part of a previous study. Details about the in vivo experiment, sample collection and processing are available at [29].

siRNAs and Transfection Methods
Reverse siRNA transfection was performed at the time of seeding at a final 20 nM concentration for a total of 48 h, unless otherwise indicated. siRNA sequences for AMBRA1 are custom designed, as reported in [29]. Negative control cells (siScr) were transfected with the MISSION® siRNA Universal Negative Control #1 (Sigma-Aldrich; cat# SIC001). Overexpression of plasmid constructs was carried for the times indicated in the specific experiments after cells had been reversely transfected for 24 h with siAMBRA#2, which was specifically designed to target the 5′-UTR region of *AMBRA1* in order to exclude effect(s) (i) of the siRNA on expression of AMBRA1 plasmid constructs and (ii) of endogenous AMBRA1 in the downstream applications (Fig. 3B). The sequence coding for wild-type AMBRA1 (WT) (UniProt ID: Q9C0C7-1) was cloned in either pcDNA™3.1 Mammalian Expression (ThermoFisher Scientific; cat# V79020) or pcDNA™3.1/myc-His A, B, & C Mammalian Expression (ThermoFisher Scientific; cat# V80020) vectors. The coding sequences were amplified by PCR and cloned in the acceptor vector by means of EcoRI and NotI restriction sites. AMBRA1 mutants (P63S, S90F, T97I, L110F, S142F, A157V, P170S) were generated by site-directed mutagenesis using AMBRA1 as template and custom designed primers. The list of mutants of the N-terminal region of AMBRA1 on which the in vitro validation has been performed does not include all the point mutations with REVEL ≥ 0.4 shown in Supplemental table S2, as DNA constructs were generated on a previous version of the mutation plot dated 22 May 2018. All transfections were performed using Lipofectamine™ 2000 Transfection Reagent (ThermoFisher Scientific; cat# 11668-019), and manufacturer's instructions were followed.

Protein expression analysis
At the time of collection, cells were washed in Phosphate Buffer Solution (PBS, ThermoFisher Scientific; cat# 14190144), mechanically detached and centrifuged at $1200 \times g$ for 5 min at 4 °C and cell pellets processed and previously described [29]. Protein concentration of supernatants was determined by the Lowry's method. For soluble/insoluble analysis, cell debris (insoluble fraction) was washed three times in RIPA buffer followed by centrifugation at $10,000 \times g$ for 5 min at 4 °C. Both soluble and insoluble fractions were denatured in NuPAGE™ LDS Sample Buffer (4X) (Thermo-Fisher Scientific; cat# NP0007) supplemented with NuPAGE™ Sample Reducing Agent (10X) (ThermoFisher Scientific; cat# NP0004) followed by incubation at 100 °C for 5 min. Protein extracts were separated by SDS-PAGE using Criterion™ TGX™ Precast Gels (Bio-Rad Laboratories; cat# 567-8084) and blotted onto PVDF membranes (Bio-Rad Laboratories; cat# 10026933) using a Trans-Blot® Turbo™ Transfer System (Bio-Rad Laboratories). Primary antibodies used are as follows:

Target	Dilution	Distributor	Cat#	RRID:
Actin	1:40,000	Novus Biologicals	NB600-501	AB_10077656
AMBRA1	1:1000	Santa Cruz Biotechnology	sc-398204	AB_2861324
AMBRA1	1:1000	Merck-Millipore	ABC131	AB_2636939
AMBRA1	1:1000	Sigma-Aldrich	PRS4557	AB_1848744
c-myc	1:1000	Santa Cruz Biotechnology	sc-40	AB_627268
CASP-3	1:1000	Cell Signaling Technology	9662S	AB_331439
CDH1	1:1000	Cell Signaling Technology	3195S	AB_2291471
CDH2	1:1000	Cell Signaling Technology	13116S	AB_2687616
Cyclin D1	1:2000	Abcam	ab16663	AB_443423
FAK1	1:1000	Cell Signaling Technology	13009	AB_2798086

Table continued

Target	Dilution	Distributor	Cat#	RRID:
LC3	1:2500	Cell Signaling Technology	3868S	AB_2137707
pFAK-Y397	1:1000	Cell Signaling Technology	8556	AB_10891442
pSRC-Y416	1:1000	Cell Signaling Technology	6943S	AB_10013641
SNAI1	1:1000	Cell Signaling Technology	3879	AB_2255011
SRC	1:1000	Cell Signaling Technology	2123S	AB_2106047
Ubiquitin (Ub)	1:1000	Merck-Millipore	05-1308	AB_1587580
VIM	1:1000	Cell Signaling Technology	5741	AB_10695459

Images were captured with a ChemiDoc™ MP System (Bio-Rad Laboratories; cat# 1708-280) provided with the Image Lab 6.0.1 Software (Bio-Rad Laboratories). Densitometry analyses were carried out using the ImageJ software (1.52.q) (University of Wisconsin; RRID:SCR_003070). Full length uncropped original western blots are provided and available in the Supplemental Material file.

RNA isolation, reverse transcription, and quantitative RT-PCR
Total RNA was isolated and reverse transcription were performed as previously described [29]. cDNA was diluted three times and mRNA expression levels detected by PowerUp™ SYBR™ Green Master Mix (ThermoFisher Scientific; cat# A25742), according to the instructions, on a ViiA 7 Real-Time PCR System v1.3 (Applied Biosystems). All reactions were run in triplicate and mRNA levels expressed as fold change (relative to control) after normalization to the internal housekeeping L34. The specific primer pairs were custom designed and tested with Primer-BLAST (NCBI; RRID:SCR_003095). Primers used were obtained from TAG Copenhagen A/S (Copenhagen, DK) and are as follows: L34: FW: 5′- GGC CCT GCT GAC ATG TTT CTT -3′, RV: 5′- GTC CCG AAC CCC TGG TAA TAG A -3′; AMBRA1: FW: 5′- AAC CCT CCA CTG CGA GTT GA -3′, RV: 5′- TCT ACC TGT TCC GTG GTT CTC -3′; FN1: FW: 5′- CGA CAC ATT CCA CAA GCG TC -3′, RV: 5′- CAT TGG TCG ACG GGA TCA CA -3′; VIM: FW: 5′- GAC GCC ATC AAC ACC GAG TT -3′, RV: 5′- CTT TGT CGT TGG TTA GCT GGT -3′; CCND1: FW: 5′- GAT CAA GTG TGA CCC GGA CT-3′, RV: 5′- CTT GGG GTC CAT GTT CTG CT-3′.

Wound healing assay
Twenty-four hours after re-expression of the plasmid constructs, 25,000 cells were seeded in each of the two wells of silicone inserts with a defined gap of 500 μm (ibidi®; cat# 80209) in six-well plates. After 16 h, the inserts were removed and wound closure followed at the times indicated. Migrating cells were imaged with an IX71 inverted microscope (Olympus) provided with a CellSens Imaging Software 2 (Olympus; RRID:SCR_016238). The area of wound closure was calculated using ImageJ with respect to the initial area (T_0) and expressed as percentage of wound healing at the time points indicated. In the representative pictures, the white and yellow lines outline the edge of the wound at T_0 and at 24 h, respectively.

Cell proliferation
Twenty-four hours after re-expression of the plasmid constructs, 10,000 cells were seeded in 12-well plates. After 24 h and 48 h, cells were washed with PBS, fixed-and-stained with a 0.025% (w/v) Crystal violet (Sigma-Aldrich; cat# C6158) solution in 20% (v/v) MeOH on ice for 15 min. After washing with ddH$_2$O, plates were air-dried and pictures taken with an IX71 inverted microscope (Olympus) provided with a CellSens Imaging Software 2 (Olympus). For quantitation, Crystal violet was eluted with 100% MeOH and absorbance measured at 595 nm by a VICTOR Multilabel Plate Reader (PerkinElmer). Data are expressed as fold change with respect to absorbance of control sample (WT at 24 h).

Cell viability
Twenty-four hours after re-expression of the plasmid constructs, 7500 cells were seeded in 96-well plates and cell viability measured at the times indicated using the Cell Counting Kit-8 (Dojindo; cat# CK04-11) at 450 nm using a VICTOR Multilabel Plate Reader (PerkinElmer) after 2 h of incubation, according to the manufacturer's instructions. Data are shown as fold change of viable cells with respect to control cells (WT at 24 h).

Statistical analysis
Ordinary one-way ANOVA was used for densitometry and RT-qPCR analyses. Two-way ANOVA was used for wound healing, cell proliferation and viability assays. All ANOVA tests were corrected using the Bonferroni multiple comparison test and statistical values calculated in function of a control case. GraphPad/Prism9 (version 9.2.0) (RRID:SCR_002798) was used for plotting graphs and to perform statistical analysis. Data are presented as means ± SEM or SD, as indicated in the figure legends, and significance was designated as follows: *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$; ns, not significant. Source data are provided within this paper.

Structured regions of AMBRA1 according to AlphaFold
We have downloaded the Alphafold [55] model for human AMBRA1 from the EMBL-EBI Alphafold Protein Structure Database [50], entry Q9C0C7. Visual inspection of the model showed a major structural feature for this model—a β-propeller folded domain spanning regions ~41–203 and ~857–1040 of AMBRA1. The AlphaFold prediction was deemed to be confident (pLDDT > 70) for the first stretch of residues and for most part of the second, with short stretches of residues at lower confidence which correspond to short solvent-exposed loops and are thus likely to be disordered. AlphaFold also predicts the N-terminus of AMBRA1 to be structured as two consecutive alpha helices, one with low confidence (residues 7–19, most of them with pLDDT scores between 50 and 70) and one at high confidence (residues 25–40, pLDDT > 70).

Free energy calculations
We trimmed the structure keeping only residue stretches corresponding to the predicted structured regions of AMBRA1 (residues 1–200 and 850–1040). We then used the MutateX pipeline [46] saturation scan protocol with FoldX 5.0 [47] to run a complete mutational scan of the resulting structure, predicting the changes of folding free energy upon mutation for the substitution of each amino acid to each natural variant for a total of 7820 data points. For each data point we considered the average difference in free energy between wild-type and mutant variant over five independent FoldX runs. The MutateX protocol includes both a repair step, in which the structure is optimized using FoldX, and generation of mutant variant structures together with folding free energy estimation.

Sequence alignment
We have obtained a protein multiple sequence alignment between different AMBRA1 orthologs using Clustal Omega [56], using the protein sequences corresponding to the main protein isoform of AMBRA1 of human, chimpanzee, mouse, rat, bovine, Xenopus and zebrafish (ambra1a for the latter).

DATA AVAILABILITY
Results from AMBRA1 Cancermuts runs are available on our GitHub repository (https://github.com/ELELAB/cancermuts/tree/master/data_case_study).

REFERENCES
1. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N. Engl J Med. 2016;375:1109–12.
2. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45:1113–20.
3. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. 2019;47:D941–7.
4. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.
5. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1.

6. Andrades R, Recamonde-Mendoza M. Machine learning methods for prediction of cancer driver genes: a survey paper. Brief Bioinform. 2022;23:1–19.

7. Rogers MF, Gaunt TR, Campbell C. Prediction of driver variants in the cancer genome via machine learning methodologies. Brief Bioinform. 2021;22:1–11.

8. Poulos RC, Wong JWH. Finding cancer driver mutations in the era of big data research. Biophys Rev. 2019;11:21–9.

9. Raphael BJ, Dobson JR, Oesper L, Vandin F. Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. Genome Med. 2014;6:5.

10. Pejavar V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun. 2020;11:5918.

11. Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, et al. CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics. 2013;29:647–8.

12. Ainscough BJ, Griffith M, Coffman AC, Wagner AH, Kunisaki J, Choudhary MN, et al. DoCM: a database of curated mutations in cancer. Nat Methods. 2016;13:806–7.

13. Cario CL, Witte JS. Orchid: a novel management, annotation and machine learning framework for analyzing cancer mutations. Bioinformatics. 2018;34:936–42.

14. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. Hum Mutat. 2015;36:E2423–9.

15. Fas BA, Maiani E, Sora V, Kumar M, Mashkoor M, Lambrughi M, et al. The conformational and mutational landscape of the ubiquitin-like marker for autophagosome formation in cancer. Autophagy. 2021;17:2818–41.

16. Nygaard M, Terkelsen T, Vidas Olsen A, Sora V, Salamanca Viloria J, Rizza F, et al. The mutational landscape of the oncogenic MZF1 SCAN domain in cancer. Front Mol Biosci. 2016;3:78.

17. Konig SM, Rissler V, Terkelsen T, Lambrughi M, Papaleo E. Alterations of the interactome of Bcl-2 proteins in breast cancer at the transcriptional, mutational and structural level. PLoS Comput Biol. 2019;15:e1007485.

18. Abildgaard AB, Stein A, Nielsen SV, Schultz-Knudsen K, Papaleo E, Shrikhande A. Computational and cellular studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome. Elife. 2019;8:1–28.

19. Nielsen SV, Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG, et al. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. PLoS Genet. 2017;13:e1006739.

20. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? J Mol Biol. 2019;431:2197–212.

21. Ponzoni L, Bahar I. Structural dynamics is a determinant of the functional significance of missense variants. Proc Natl Acad Sci USA. 2018;115:4164–9.

22. Fimia GM, Stoykova A, Romagnoli A, Giunta L, Di Bartolomeo S, Nardacci R, et al. Ambra1 regulates autophagy and development of the nervous system. Nature 2007;447:1121–5.

23. Cianfanelli V, De Zio D, Di Bartolomeo S, Nazio F, Strappazzon F, Cecconi F. Ambra1 at a glance. J Cell Sci. 2015;128:2003–8.

24. Nazio F, Strappazzon F, Antonioli M, Bielli P, Cianfanelli V, Bordi M, et al. mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. Nat Cell Biol. 2013;15:406–16.

25. Cianfanelli V, Fuoco C, Lorente M, Salazar M, Quondamatteo F, Gherardini PF, et al. AMBRA1 links autophagy to cell proliferation and tumorigenesis by promoting c-Myc dephosphorylation and degradation. Nat Cell Biol. 2015;17:20–30.

26. Maiani E, Milletti G, Nazio F, Holdgaard SG, Bartkova J, Rizza S, et al. AMBRA1 regulates cyclin D to guard S-phase entry and genomic integrity. Nature 2021;592:799–803.

27. Simoneschi D, Rona G, Zhou N, Jeong YT, Jiang S, Milletti G, et al. CRL4(AMBRA1) is a master regulator of D-type cyclins. Nature. 2021;592:789–93.

28. Chaikovsky AC, Li C, Jeng EE, Loebell S, Lee MC, Murray CW, et al. The AMBRA1 E3 ligase adaptor regulates the stability of cyclin D. Nature 2021;592:794–8.

29. Di Leo L, Bodemeyer V, Bosisio FM, Claps G, Carretta M, Rizza S, et al. Loss of Ambra1 promotes melanoma growth and invasion. Nat Commun. 2021;12:2550.

30. Antonioli M, Albiero F, Nazio F, Vescovo T, Perdomo AB, Corazzari M, et al. AMBRA1 interplay with cullin E3 ubiquitin ligases regulates autophagy dynamics. Dev Cell. 2014;31:734–46.

31. Strappazzon F, Nazio F, Corrado M, Cianfanelli V, Romagnoli A, Fimia GM, et al. AMBRA1 is able to induce mitophagy via LC3 binding, regardless of PARKIN and p62/SQSTM1. Cell Death Differ. 2015;22:419–32.

32. Di Bartolomeo S, Corazzari M, Nazio F, Oliverio S, Lisi G, Antonioli M, et al. The dynamic interaction of AMBRA1 with the dynein motor complex regulates mammalian autophagy. J Cell Biol. 2010;191:155–68.

33. Strappazzon F, Vietri-Rudan M, Campello S, Nazio F, Florenzano F, Fimia GM, et al. Mitochondrial BCL-2 inhibits AMBRA1-induced autophagy. EMBO J. 2011;30:1195–208.

34. Strappazzon F, Di Rita A, Cianfanelli V, D'Orazio M, Nazio F, Fimia GM, et al. Prosurvival AMBRA1 turns into a proapoptotic BH3-like protein during mitochondrial apoptosis. Autophagy 2016;12:963–75.

35. Van Humbeeck C, Cornelissen T, Hofkens H, Mandemakers W, Gevaert K, De Strooper B, et al. Parkin interacts with Ambra1 to induce mitophagy. J Neurosci. 2011;31:10249–61.

36. Antonioli M, Pagni B, Vescovo T, Ellis R, Cosway B, Rollo F, et al. HPV sensitizes OPSCC cells to cisplatin-induced apoptosis by inhibiting autophagy through E7-mediated degradation of AMBRA1. Autophagy 2021;17:2842–55.

37. Di Rienzo M, Antonioli M, Fusco C, Liu Y, Mari M, Orhon I, et al. Autophagy induction in atrophic muscle cells requires ULK1 activation by TRIM32 through unanchored K63-linked polyubiquitin chains. Sci Adv. 2019;5:eaau8857.

38. Miki Y, Tanji K, Mori F, Tatara Y, Utsumi J, Sasaki H, et al. AMBRA1, a novel alpha-synuclein-binding protein, is implicated in the pathogenesis of multiple system atrophy. Brain Pathol. 2018;28:28–42.

39. Xia P, Wang S, Huang G, Du Y, Zhu P, Li M, et al. RNF2 is recruited by WASH to ubiquitinate AMBRA1 leading to downregulation of autophagy. Cell Res. 2014;24:943–58.

40. Di Rita A, Peschiaroli A, D Acunzo P, Strobbe D, Hu Z, Gruber J, et al. HUWE1 E3 ligase promotes PINK1/PARKIN-independent mitophagy by regulating AMBRA1 activation via IKKalpha. Nat Commun. 2018;9:3755.

41. UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:D480–9.

42. Ioannidis NM, Rothstein JH, Pejavar V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99:877–85.

43. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.

44. Kumar M, Gouw M, Michael S, Samano-Sanchez H, Pancsa R, Glavina J, et al. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 2020;48:D296–D306.

45. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43:D512–20.

46. Tiberti M, Terkelsen T, Degn K, Beltrame L, Cremers TC, da Piedade I. MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles. Brief Bioinform. 2022;23:1–16.

47. Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. Bioinformatics. 2019;35:4168–9.

48. Zhang X, Walsh R, Whiffin N, Buchan R, Midwinter W, Wilk A, et al. Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. Genet Med. 2021;23:69–79.

49. Zhao J, Bian ZC, Yee K, Chen BP, Chien S, Guan JL. Identification of transcription factor KLF8 as a downstream target of focal adhesion kinase in its regulation of cyclin D1 and cell cycle progression. Mol Cell. 2003;11:1503–15.

50. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50:D439–44.

51. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature. 2021;599:91–5.

52. Laine E, Karami Y, Carbone A. GEMME: a simple and fast global epistatic model predicting mutational effects. Mol Biol Evol. 2019;36:2604–19.

53. Riesselman AJ, Ingraham JB, Marks DS. Deep generative models of genetic variation capture the effects of mutations. Nat Methods. 2018;15:816–22.

54. Hopf TA, Ingraham JB, Poelwijk FJ, Scharfe CP, Springer M, Sander C, et al. Mutation effects predicted from sequence co-variation. Nat Biotechnol. 2017;35:128–35.

55. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

56. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50:W276–9.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS
Conception and design: MT, LDL, DDZ, and EP; Development of methodology: MT, LDL, DDZ, and EP; Acquisition of data: MT, LDL, MVV, RSK, and EP; Analysis and interpretation of data: MT, LDL, DDZ, and EP; Writing, review and/or revision of the

manuscript: MT, LDL, FC, DDZ, and EP (all authors have reviewed and edited the manuscript); Administrative, technical, or material support: MT, LDL, FC, DDZ, and EP; Study supervision: DDZ and EP.

FUNDING

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS STATEMENT

No human participants, human tissue, or animal model are involved in this study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41419-022-05318-2.

Correspondence and requests for materials should be addressed to Matteo Tiberti or Daniela De Zio.

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.