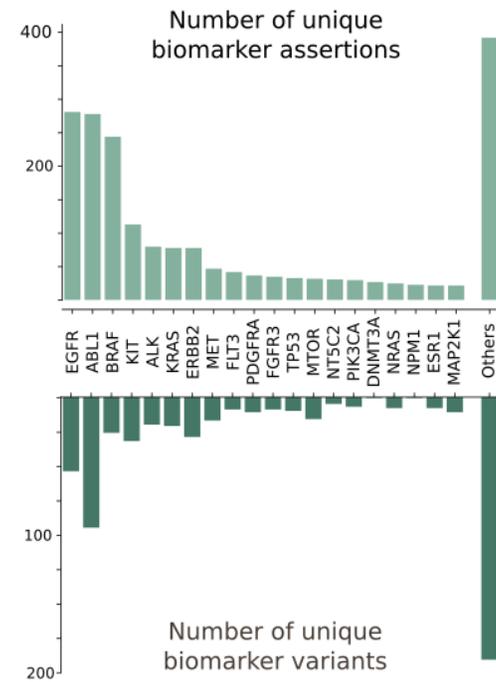
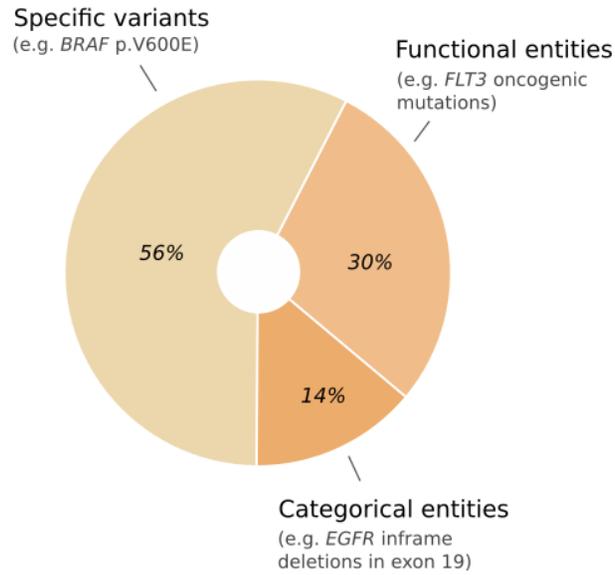


Suppl. Figure 1

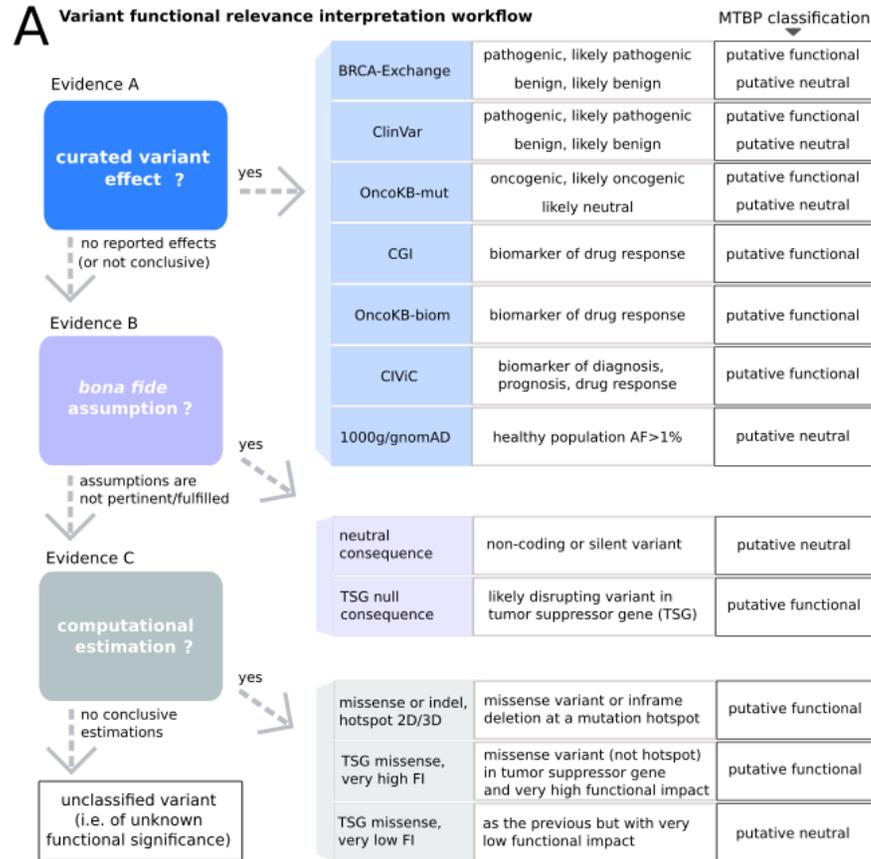
Variants reported as cancer biomarkers



Overview of gene mutations (single nucleotide variants and small indels) reported as biomarkers of cancer diagnosis, prognosis and/or drug response by three publicly available knowledgebases (CIViC, OncoKB and Cancer Genome Interpreter) at the moment of writing this manuscript. "Assertion" represents a given combination of gene mutation, biomarker effect, cancer type and

level of clinical evidence (such as "reported in early clinical trials"). Assertions supported by weak or inconclusive evidence (as provided by the knowledgebase metadata when appropriate, for example less than three stars in the CIViC evidence rating) are excluded from these results.

Suppl. Figure 2



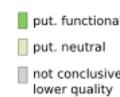
B Rationale for the strength of the evidence supporting the variant functional relevance

Evidence A	An established variant effect is considered <u>very strong</u> or <u>strong</u> supporting evidence (depending on the evidence source)
Evidence B	Only <i>bona fide</i> biological assumptions considered <u>very strong</u> or <u>strong</u> are used (see panel C for further details)
Evidence C	Hotspots detected by statistical methods modeling genomic mutational processes as well as functional impact prediction with the selected metrics (see panel D for benchmarked performance) are both considered <u>strong</u> evidence

C Knowledgebases variant aggregation

(excluding population genetics data)

554,911 records
364,009 unique variants

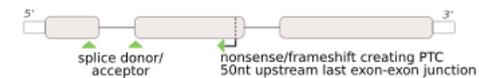


Number of variants curated by *N* knowledgebases

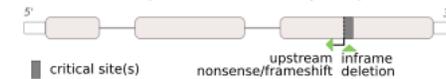


D Disrupting variants in tumor suppressors

Protein product is likely not produced or disrupted:



Critical site(s) for protein function is likely disrupted:

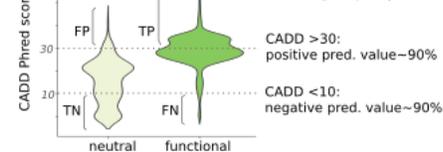


More than 10% of the protein sequence disrupted:

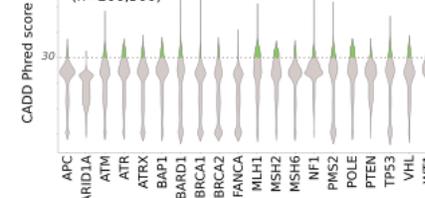


E Functional impact score benchmarking

Missense variants with curated effect in TSGs (n=2,587)



Simulation of possible missense variants in TSGs (n=206,500)



3Panel A: The MTBP classifies a given cancer gene variant as (putative) functionally relevant or neutral according to three distinct sources of evidence (named as A, B and C here), or as unclassified (i.e. of unknown functional significance) if none of these criteria are fulfilled. Note that the knowledgebases listed here are those integrated at the moment of writing this manuscript, but this may be subject to changes depending on evolving needs and preferences.

Panel B: Criteria employed to support the variant functional classification are considered to provide strong (>0.9 certainty) or very strong (>0.99 certainty) evidence as extrapolated from the work in variant pathogenicity classification (see S. Tavtigian et al, Genet. Med. 2018), according to the rationale described in the table.

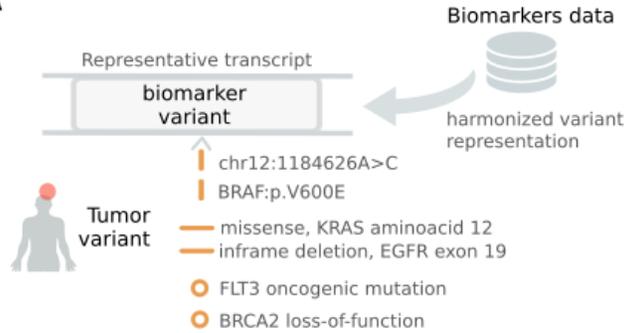
Panel C: The MTBP aggregates more than half a million of assertions curated in several knowledgebases (without counting those provided by genetics population data) as supporting evidence of type A at the moment of manuscript writing. Most of these assertions are supported by weaker data (for example not well powered studies, as retrieved from the respective knowledgebase metadata as appropriate) or do not contain conclusive information, and are therefore not considered for classifying the variant. As expected by the different scopes of each knowledgebase and the long tail of lowly recurrent mutations, only a minority of the variants appear curated in more than one knowledgebase, which stresses the value of aggregating their content for a more comprehensive functional annotation.

Panel D: Graphical summary of some of the criteria used for assuming that a variant with null consequence type is disrupting the function of a given tumor suppressor gene (included as part of the evidence set of type B). These are largely based on established rules to identify loss-of-function variants in Mendelian genes (see S. Richards et al., Genet. Med. 2015), with subsequent refinements (see A. Abou Tayoun et al, Hum. Mutat. 2018) exemplified here. Top: canonical splice site disruptions and variants leading to a premature stop codon (likely) triggering nonsense-mediated decay mechanisms are considered very strong criteria (so-called somePVS1). Middle: variants truncating/disrupting protein regions that are crucial for the tumor suppressor function is strong supporting criteria (PVS1_Strong). Of note, the identification of these crucial protein regions is refined by analyzing location and consequence type of known loss-of-function variants, as supported by the knowledgebases aggregated by the MTBP. Bottom: variant truncating/disrupting more than 10% of the wildtype tumor suppressor protein is strong supporting criteria (PVS1_Strong).

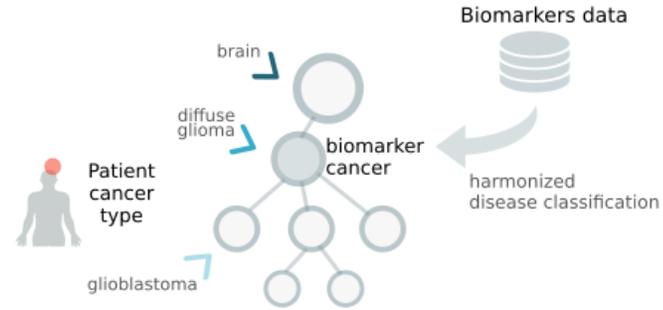
Panel E: The lowest level evidence used by the MTBP to estimate a given variant effect is based on bioinformatics metrics. For variants that are not located within (somatic) mutation hotspots, we decided to use the Combined Annotation Dependent Depletion (CADD) to estimate the functional relevance of missense mutations in tumor suppressor genes (TSGs), as functional impact predictions perform worse in other scenarios (data not shown). Importantly, the CADD score thresholds were selected after benchmarking the method's performance for all the missense variants observed in TSGs with known effect (as retrieved from the aggregated knowledgebases content). In detail, missense variants with very high (>30) and very low (<10) Phred CADD scores exhibit 90% of true putative functional and neutral calls, respectively (upper violin plot), as required for being considered strong criteria (see S. Tavtigian et al., Genet. Med. 2018). Of note, we simulated *in silico* all the nucleotide changes leading to missense variants that can occur in some of the most recurrently mutated TSGs, and we observed that most of these variants remain unclassified by using these CADD thresholds (lower violin plot), which underscores the stringency of the criteria. FP=false positives; TP=true positives; TN=true negatives; FN=false negatives.

Suppl. Figure 3

A Variant matching: tumor sample versus cancer biomarker



B Cancer type matching: tumor sample versus cancer biomarker



C Actionability tiers of the cancer biomarker in the tumor sample

AMP/ASCO/CAP classification	ESMO-ESCAT classification	MTBP biomarker classification	Sample versus biomarker match		Current biomarker clinical relevance		
			Mutation	Cancer type	CGI	CIViC	OncoKB
Level IA Approved guidelines Professional recommendations	Ready for use	I. Ready for use I. Ready for use- pancancer	Yes	Yes	approved	A	Level 1
	IA clinical benefit, randomized trial				recommended		Level 2
	IB clinical benefit, non-randomized trial						
Level IB Well-powered studies with expert consensus	Investigational	II. Investigational II. Investigat.- case report	Yes	Yes	late clinical trials early clinical trials	B	Level 3
	IIA clinical benefit, retrospective evaluation				case report		C
Level IIC Small studies with some consensus Cancer repurposing	Hypothetical - repurposing	III. Cancer repurposing III. Mutation repurposing	Yes	No	see tiers I/II	see tiers I/II	see tiers I/II
	IIIA tier I/II for another cancer						
	IIIB tier I/II for equivalent mutation						
Level IID Pre-clinical data	Hypothetical - no clinical data	IV. Pre-clinical IV. in silico	Yes	Yes	pre-clinical data	D	Level 4
	IVA pre-clinical data				E		

Biomarker type than can be included in the MTBP classification:
 ▲ Drug increased sensitivity/response
 ▼ Drug decreased sensitivity/resistance
 ◯ Drug negative evidences
 ◐ Diagnostic or prognostic

Panel A: Tumor variants are matched with cancer biomarkers reported as a specific nucleotide or protein change (see vertical line examples), a categorical genomic definition (see horizontal line examples) or a functional entity (which are matched according to the variant's MTBP functional interpretation; see circle examples).

Panel B: Patient and biomarker cancer types are compared taking into account the disease hierarchy (as represented by the cancer taxonomy tree). In detail, the cancer biomarker is matched if reported for the patient's tumor type or a subtype thereof (medium and low-lighted blue arrows, respectively).

Panel C: Genomic cancer biomarkers (diagnosis, prognosis and drug response) found in the tumor are ranked following the European Society for Medical Oncology (ESMO) Clinical Actionability of Molecular Targets (ESCAT) scale. We introduced two minor modifications as displayed here due to (i) the lack of structured information in the knowledgebases to infer the trial design details; and (ii) the use of biomarkers reported by clinical observations only supported by case reports. An approximate equivalence of the tumor actionability categories proposed by the Association for Molecular Pathology/American Society of Clinical Oncology/College of American Pathologists (AMP/ASCO/CAP) and the ESMO-ESCAT is shown here (left). The color table summarizes how the MTBP biomarker category is defined according to the coincidence of the (i) variant and (ii) cancer type reported for the patient's tumor *versus* biomarker; and (iii) the current clinical evidence of the biomarker effect (as classified by each respective knowledgebase).

Suppl. Figure 4

A Gene variants analysis in the public MTBP website

The screenshot shows the MTBP website interface for analyzing single nucleotide variants and indels. It features a sidebar with navigation options: 'Your reports', 'Analyse', 'Account', and 'FAQ'. The main content area is titled 'Analyse single nucleotide variants and indels' and is divided into three numbered steps:

- 1 Provide an analysis identifier**: A text input field containing 'coadread_001'.
- 2 Provide the cancer type**: A dropdown menu with a list of cancer types, including 'Bladder/Urinary Tract', 'Blood', 'Bone', 'Bowel', 'Anal Gland Adenocarcinoma', 'Anal Squamous Cell Carcinoma', 'Anorectal Mucosal Melanoma', 'Appendiceal Adenocarcinoma', 'Colorectal Adenocarcinoma' (highlighted), and 'Gastrointestinal Neuroendocrine Tumors'.
- 3 Provide gene variants and run**: Two options are provided:
 - > Upload a VCF file: A 'Choose File' button with 'No file chosen' and a 'Run' button.
 - OR
 - > Introduce variants in HGVS format: A text input field containing 'KRAS:p.K12C', 'CHEK1:c.207A>C', 'FAM175A:c.908delA', and 'chr3:g.52437215C>T', with a 'Run' button.

B Public MTBP access statistics

(April 2019-January 2021)

~1,400 unique users
~20,00 website access



Panel A: The public MTBP website provides a general framework to annotate the functional and predictive relevance of gene mutations (only single nucleotide variants and small indels) uploaded to the system. This resource is open for research purposes only and does not incorporate *ad-hoc* interpretation nuances (such as supporting evidence required for considering functional events in actionable genes) or *in-house* actionability flags (such as eligibility for Cancer Core Europe ongoing clinical trials).

Panel B: Public MTBP website activity from the date of its release until the moment of writing this manuscript.

Appendix 1. List of genes evaluated for germline variants with inherited increased cancer risk. These are the genes recommended by the American College of Medical Genetics and Genomics (ACMG SF v2.0 update, see S Kalia ScM et al., *Genetics in Medicine* 2017), plus some additions (marked with an asterisk) agreed to be clinically actionable by the Cancer Core Europe Genetic Counseling Task Force. Of note, the MTBP flags genetic counseling alerts only if the germline variants observed in these genes are estimated to be loss-of-function based on well-curated evidence and/or *bona fide* biological assumptions (see Suppl. Figure 2), regardless of patient clinical information.

APC, ATM*, BMPR1A, BRCA1, BRCA2, BRIP1*, CDH1*, CHEK2, MEN1, MLH1, MSH2, MSH6, MUTYH^(a), NF2, PALB2*, PMS2, PTEN, RAD51C*, RAD51D*, RB1, RET, SDHB, SDHC, SDHD, SMAD3, SMAD4, STK11, TGFBR1, TGFBR2, TP53, TSC1, TSC2, VHL, WT1

^(a) *only reported for homozygous or compound heterozygous MUTYH mutations*