The context-specific role of germline pathogenicity in tumorigenesis

Preethi Srinivasan1,2,7, Chaitanya Bandlamudi1,2,7, Philip Jonsson2, Yelena Kemel3, Shweta S. Chavan2, Allison L. Richards2, Alexander V. Penson2,4, Craig M. Bielski4,5, Christopher Fong2,5, Aijazuddin Syed1, Gowtham Jayakumar1, Meera Prasad1, Jason Hwee1, Selcuk Onur Sumer1,2, Ino de Bruijn2, Xiang Li2, JianJiong Gao2,5, Nikolaus Schultz2,4,5, Roy Cambria6, Jesse Galle6, Semanti Mukherjee3,7, Joseph Vijai3,7, Karen A. Cadoo3,7, Maria I. Carlo3,7, Michael F. Walsh3,7, Diana Mandelker1, Ozge Ceyhan-Birsoy1, Jinru Shia1, Ahmet Zehir1, Marc Ladanyi1,4, David M. Hyman7,9, Lijing Zhang1,6, Kenneth Offit3,7, Mark E. Robson3,7, David B. Solit2,4,7, Zsofia K. Stadler3,7, Michael F. Berger2,4, and Barry S. Taylor2,4,5,9

Human cancers arise from environmental, heritable and somatic factors, but how these mechanisms interact in tumorigenesis is poorly understood. Studying 17,152 prospectively sequenced patients with cancer, we identified pathogenic germline variants in cancer predisposition genes, and assessed their zygosity and co-occurring somatic alterations in the concomitant tumors. Two major routes to tumorigenesis were apparent. In carriers of pathogenic germline variants in high-penetrance genes (5.1% overall), lineage-dependent patterns of biallelic inactivation led to tumors exhibiting mechanism-specific somatic phenotypes and fewer additional somatic oncogenic drivers. Nevertheless, 27% of cancers in these patients, and most tumors in patients with pathogenic germline variants in lower-penetrance genes, lacked particular hallmarks of tumorigenesis associated with the germline allele. The dependence of tumors on pathogenic germline variants is variable and often dictated by both penetrance and lineage, a finding with implications for clinical management.
systematically characterize the role of pathogenic germline alleles on tumorigenesis, and we describe a new online resource (https://www.signaldb.org) to facilitate broader investigation and improved interpretation of germline variants by the scientific and clinical community.

**Results**

**Germline pathogenic variants in common and rare cancers.** To understand the role of germline pathogenicity in tumorigenesis, we analyzed prospectively acquired sequencing data from 17,152 patients with cancer diagnoses spanning 55 broad cancer types and 413 histological subtypes (Supplementary Table 1). Germline blood and matched tumor tissue DNA were sequenced using Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT), a Food and Drug Administration (FDA)-authorized clinical sequencing assay encompassing up to 468 cancer-associated genes, including 84 well-established cancer predisposition genes (Methods). Germline variant calling was performed based on a sequence analysis pipeline validated for clinical use in a Clinical Laboratory Improvement Amendments (CLIA)-compliant laboratory. To identify pathogenic germline variants related to cancer, we developed a machine learning-based framework that integrated mutation type, functional gene class, population allele frequencies and orthogonal in silico tools to predict functional impact (Extended Data Fig. 1). We then trained this classifier (cross-validated precision and sensitivity of 94 ± 6% and 89 ± 7%, respectively) on a cohort of 6,009 patients with cancer whose germline data were prospectively curated by expert clinical molecular geneticists using American College of Medical Genetics guidelines for clinical interpretation (Extended Data Fig. 1; Supplementary Note).

We identified 1,499 unique (3,330 total) predicted pathogenic or likely pathogenic germline mutations and copy-number variants (for brevity, hereafter referred to as ‘pathogenic’ variants) in 183 genes. We then stratified the pathogenic variants into three penetrance categories (high, moderate and low) using previously established community standards and manual review (Supplementary Note, Fig. 1a; Extended Data Fig. 1 and Supplementary Table 2). All additional pathogenic variants were designated as uncertain penetrance, 44% of which occurred in known cancer predisposition genes (Supplementary Tables 2 and 4). The overall frequency of germline pathogenicity reflected, in part, the composition of ancestry in our patient cohort, which we estimated to comprise 18% of patients of Ashkenazi Jewish ancestry, in whom a higher frequency of founder pathogenic germline variants has been reported (Extended Data Fig. 2 and Supplementary Note). Overall, we identified a pathogenic germline variant in genes conferring markedly elevated risk for cancer incidence (high and moderate penetrance) in 7.8% of patients (Supplementary Tables 3 and 4). Variants with marginally elevated risk for cancer (low penetrance) were identified in 2.9% of patients and exclusively comprised the Ashkenazi founder variant APC p.I1307K and heterozygous variants in MUTYH. The frequency of the APC p.I1307K variant among the Ashkenazi Jewish patients (9.1%) is consistent with previous reports. Notably, the frequency of MUTYH heterozygous carriers is identical among patients with colorectal and other cancers (1.5%), and consistent with their prevalence in the general population, further reflecting the low penetrance effects of these variants. A higher proportion of patients of Ashkenazi ancestry were carriers of high/moderate penetrance variants (11.4%) compared with patients of European (6.9%), African (5.9%) or Asian (5.2%) ancestries.

We next performed ancestry-adjusted association testing to determine the cancer types, at the time of genetic analysis, that are significantly enriched for pathogenic germline variants in a given gene. We identified the majority of known associations between cancer susceptibility genes and tumor types, as well as emerging associations even in well-studied genes such as BRCA2 (Fig. 1b and Supplementary Table 5). Nevertheless, 47% of all cancers with high-penetrance pathogenic variants involved cancer types in which no prior association has been previously established, raising the possibility that certain tumors may have formed independently of the underlying pathogenic germline allele.

**Selective pressure for somatic zygosity changes.** Loss of the wild-type (WT) allele is a critical event for tumorigenesis in carriers of pathogenic variants in the majority of tumor-suppressor cancer predisposition genes. We therefore assessed tumor-specific biallelic inactivation via somatic copy-number loss of heterozygosity (LOH) or second somatic mutations as a measure of the role of pathogenic alleles on tumor formation and progression. We considered 57 out of 176 tumor-suppressor genes in which pathogenic variants were identified in ≥5 patients. In each carrier, we determined tumor-specific zygosity changes by integrating high-precision mutant allele frequencies with purity- and ploidy-corrected, allele-specific copy-number data (Methods). To account for site-specific background rates of DNA copy-number alterations at loci of interest, we then compared the rates of biallelic inactivation in tumors harboring pathogenic germline alleles against a background distribution of tumors harboring germline variants in the same genes predicted to have no effect on fitness (that is, benign variants and variants of unknown significance (VUSs), Methods). Overall, the rate of somatic biallelic inactivation was 40% among all affected carriers, significantly greater than the rate of second hits affecting predicted benign variants in the same genes (22.1%; \(P = 5.7 \times 10^{-16}\)). However, this rate varied widely among high and moderate cancer predisposition genes, ranging from 89% in MLH1 to 18% in NBN, suggesting that additional molecular and clinical contexts may drive this variability (Fig. 1b). Altogether, 91% of all patients harboring biallelic inactivation acquired the second hit via LOH, with notably higher utilization of somatic mutation as a second hit in MSH2, NFI, APC and CDH1 (Extended Data Fig. 3). Although we cannot exclude the possibility of additional cryptic noncoding alterations or promoter hypermethylation as further contributors to biallelic inactivation, multimodality genomic analysis of The Cancer Genome Atlas (TCGA) data indicates that epigenetic silencing arises largely mutually exclusively with germline variants in key cancer predisposition genes, and accounts for a negligible fraction of biallelic inactivation in germline carriers pan-cancer (Extended Data Fig. 4).

**Zygosity is shaped by penetrance and tumor lineage.** Given the prevalence and variability of biallelic inactivation affecting germline pathogenic alleles, we sought to establish the determinants of zygosity changes in the tumors of germline carriers. We first considered the effect of different levels of penetrance. The rate of tumor-specific biallelic inactivation was greatest for high-penetrance genes (65% for pathogenic variants versus 23% for benign variants, \(P = 2.8 \times 10^{-125}\)) and decreased for the lower penetrance categories, culminating in no significant difference between pathogenic alleles and benign variants in genes of low and uncertain penetrance (Fig. 2a). This trend remained even when using different classes of ‘benign’ variants (including common variants) and for both the primary site and the metastatic sites of disease (Extended Data Fig. 5a,b).

Beyond the effect of penetrance, we also reasoned that tumor lineage would influence the selective pressure for biallelic inactivation, which we recently demonstrated for mutant BRCA1/2 (ref. 10). We grouped cancer types according to whether or not they were associated with increased incidence of pathogenic germline variants in each gene, from either our own ancestry-adjusted association testing or prior population-based studies (Methods; Supplementary Table 5). The rate of biallelic inactivation in high-penetrance genes for germline carriers with associated tumor types was 85% compared...
Fig. 1 | Germline pathogenicity in prospectively characterized advanced cancers. a, Study schema describing the identification and pathogenicity assessment of germline carriers and the number and percentage of affected patients; 3,158 pathogenic germline variants and 172 germline copy-number deletions were identified and 302 patients harbored more than one pathogenic germline variant and were annotated according to the higher-penetrance category. b, Landscape of pathogenic germline variants by penetrance and cancer type. The size of the circle denotes prevalence of each gene broken down by biallelic and heterozygous rates. GIST, gastrointestinal stromal tumor; GNET, gastrointestinal neuroectodermal tumor; NSCLC, nonsmall cell lung carcinoma. The bar chart at the top denotes rate of biallelic inactivation. The color of the circles corresponds to the rate of biallelic inactivation of the gene within the cancer type. The diamonds denote singletons. The bar chart on the right denotes overall prevalence of each gene broken down by biallelic and heterozygous rates. GIST, gastrointestinal stromal tumor; GNET, gastrointestinal neuroectodermal tumor; NSCLC, nonsmall cell lung carcinoma.

These broader patterns of penetrance- and lineage-dependent enrichment for biallelic inactivation were evident at the level of individual genes as well. All evaluable high-penetrance tumor-suppressor genes had a higher rate of biallelic inactivation in carriers of pathogenic germline alleles compared with benign variants (Fig. 2c and Supplementary Table 6). However, not all canonical tumor-suppressor genes reached statistical significance for biallelic inactivation pan-cancer, such as TP53 and APC. These largely represented genes with high somatic mutation rates across all tumors irrespective of their germline status, therefore affecting our power to detect enrichment in germline carriers. Others include those genes for which the rate of biallelic inactivation was higher when considering only associated cancer types (Extended Data Fig. 5d). For many of the high-penetrance genes with too few total mutations to be significant pan-cancer (CDH1, CDKN2A, FLCN, TSC2 and PMS2), the effect of lineage was still evident. For example, two of two renal cell cancers had somatic biallelic inactivation of germline pathogenic variants in FLCN that are associated with Birt–Hogg–Dubé syndrome and an increased risk of renal cancer25. Likewise, biallelic inactivation was evident in three of four gastric cancers with germline CDH1 mutations, which are associated with hereditary diffuse gastric cancer syndrome. However, in other...
We next sought to determine the broader clinical and molecular consequences of pathogenic germline alleles and their associated somatic changes in zygosity. The enrichment for biallelic inactivation of high-penetrance genes suggests their importance as key drivers of tumorigenesis in affected patients, so we hypothesized that such patients may exhibit different clinical, genetic and evolutionary phenotypes driven by a distinct pathogenesis. Indeed, carriers of pathogenic variants in high-penetration geneticinactivation had a significantly younger age of disease onset compared with germline WT patients with cancer ($P=6\times10^{-14}$; Fig. 3a), a difference that was restricted to cancer types associated with increased incidence in the corresponding genes ($P=8\times10^{-13}$; Extended Data Fig. 7).

We also hypothesized that the corresponding tumors in these patients would exhibit unique somatic phenotypes beyond the affected gene. Using decomposition of the mutational load of affected cancers, we characterized somatic mutational signatures associated with germline carriers. We observed that BRCA1 and BRCA2 pathogenic germline alleles were associated with signature 3, a signature of homologous recombination deficiency, germline MUTYH carriers had tumors with signature 18, a signature of
8-oxoguanine-associated mutagenesis and carriers of germline mismatch repair (MMR) defects (MLH1, MSH2, MSH6 and PMS2) had tumors with signatures of MMR deficiency (signatures 6, 15, 20, 21 and 26) (Extended Data Fig. 8) (36, 37). The mutational spectra of cancers in high-penetration germline carrier cancers therefore directly reflected a phenotypic path to transformation driven by the germline allele. These somatic phenotypes were, however, largely driven by biallelic inactivation of the underlying gene, whereas those tumors that retained heterozygosity had little evidence of haploinsufficiency mediating the same somatic phenotypes38.

Examining the acquisition of additional somatic drivers, we explored the evolution of cancers in germline carriers and compared them with sporadic tumors. We found that the tumors in germline carriers of high-penetration pathogenic variants that underwent somatic biallelic inactivation were significantly less likely to harbor a similar proportion of somatic oncogenic drivers in genes such as KRAS as sporadic cancers. Finally, the average number of somatic drivers including both gain- and loss-of-function alterations was lowest in tumors with biallelic inactivation of high-penetration genes (P = 8.3 x 10^-7, multivariate test adjusted for tumor type and sample type; Fig. 3d). These tumors were also depleted for the hallmarks of environmental exposures underlying sporadic cancers. For example, none of the lung cancers with biallelic inactivation in associated genes (0/4) exhibited a signature of tobacco exposure, which was common in tumors with heterozygous pathogenic variants in nonassociated high-penetration genes (16/23) (P = 0.019) and tumors in germline WT patients (891/1,480) (P = 0.025).

In summary, this depletion of somatic driver mutations, along with the aforementioned clinical and molecular differences, suggests a unique route to tumorigenesis in patients with pathogenic germline alleles in high-penetration genes that undergo biallelic inactivation. These cancers require fewer somatic oncogenic dependencies to confer a selective growth advantage, emphasizing the distinct etiological role of the germline allele in disease pathogenesis. The pathogenic germline variant is thus probably the dominant evolutionary force in these resulting cancers.

Somatic phenotypes associated with inherited MMR deficiency. To further explore the role of zygosity and lineage as key drivers of germline pathogenesis, we focused on patients with Lynch syndrome (LS) where pathogenic alleles in MMR genes predispose for the somatic phenotype of microsatellite instability (MSI), a tumor-agnostic biomarker of response to immune checkpoint
Articles NaTurE GENETIcS

Biallelic, MSS Het., MSI

Articles

allele diagnostic for LS33–35, we characterized the MMR gene status of patients with a multitude of cancer types35, our results suggest that types (Fig. 4b). Although MSI arises sporadically in germline WT of MSI were most prevalent in primary Lynch-associated cancer biallelic inactivation of the underlying MMR gene and the presence with LS inferred from the literature (Lynch associated, secondary Classifying cancer types according to their degree of association inactivation varied widely (P = 3.1 × 10−7). Altogether, somatic biallelic inactivation of the associated MMR gene was a near-obligate for the MSI phenotype, with 80% of biallelic tumors and only 3% of heterozygous tumors exhibiting MSI (odds ratio (OR) = 124, P = 9.4 × 10−12; Fig. 4a). Moreover, lineage appeared to further influence the manifestation of this phenotype. Classifying cancer types according to their degree of association with LS inferred from the literature (Lynch associated, secondary Lynch and non-Lynch; see Supplementary Note), we found that biallelic inactivation of the underlying MMR gene and the presence of MSI were most prevalent in primary Lynch-associated cancer types (Fig. 4b). Although MSI arises sporadically in germline WT patients with a multitude of cancer types34, our results suggest that in germline MMR carriers with tumor types where prior studies have demonstrated no association with LS, including breast cancer, lung cancer, melanoma and thyroid cancer, the MSI phenotype is very rare. We observed no MSI in such tumors in our cohort, despite 29% of these tumors harboring somatic biallelic inactivation of the MMR gene (P = 2.5 × 10−12).

Considering each MMR gene separately, the rate of biallelic inactivation varied widely (P = 3.1 × 10−7, Fig. 4c), ranging from 95% for MSH2 to 33% for PMS2. Although the overall proportion of MSI tumors also varied by gene altered (P = 4.9 × 10−8), the relative proportions of MSI tumors among carriers whose tumors show biallelic inactivation were not significantly different (P = 0.13). This gene-level variation in biallelic inactivation rate was largely attributable to gene-specific associations with different cancer types. We found that the proportion of primary Lynch-associated tumor types varied significantly across carriers of germline MMR alterations in different genes (P = 4.6 × 10−4; Fig. 4d). Carriers of LS mutations in MLH1 or MSH2 were far more likely to have a canonical Lynch-associated tumor compared with carriers of MSH6 or PMS2 alterations (OR = 5.68, P = 5.3 × 10−3), consistent with prior reports highlighting the lower lifetime risk of cancer in LS patients with MSH6 or PMS2 mutations compared with MLH1 or MSH2 mutations6. In addition, tumors of MSH6 carriers in LS- associated cancers had a lower rate of MSI (67%) compared with MLH1/MSH2 (95%) (P = 0.018; PMS2 excluded due to low carrier count; Extended Data Fig. 10a). Despite MSH6 tumors in these cancer types having tumor mutational burden rates comparable to the other MMR genes, we observed a markedly lower intensity of the MSI phenotype as measured by MSI sensor (P = 7 × 10−3) and a lower proportion of somatic indels in the affected tumors (P = 6 × 10−4; Extended Data Fig. 10b–d). We also observed a lower immunogenic burden in MSH6 tumors compared with MLH1/MSH2 carriers, as measured by the number of human leukocyte antigen-binding neoepitopes generated per non- synonymous mutation (Methods) in each tumor (Extended Data Fig. 10e), indicating that the gene-specific somatic phenotypic differences in MSI-positive, LS-associated cancers may further modulate the response to immune checkpoint blockade therapies.

Many germline alleles are dispensable to disease pathogenesis. Collectively, these data lead to the prediction that many cancers that arise in carriers of pathogenic germline alleles are unrelated to the germline variant. We reasoned that, in such tumors, if a germline allele were dispensable, DNA copy-number losses that typify the tumor genomes of advanced cancers could be acquired somatically, which deletes the pathogenic germline allele from the dominant clone without a detrimental effect on tumorigenesis and/or disease progression. Indeed, we found that 13.6% of all pathogenic germline alleles were deleted in their respective tumors, although this was least common in patients who had germline pathogenic alterations in higher-penetrance genes (P = 1.6 × 10−7; Cochran–Armitage test for trend; Fig. 5a). Moreover, the rate of somatic loss of the pathogenic germline allele was significantly greater in nonassociated cancer types (P = 1.5 × 10−4, Fig. 5b). These data suggest that, in some patients, either their cancers lose dependence on a pathogenic germline allele during tumor progression or, alternatively, they have a tumorigenesis that is independent of their inherited defect.
More subtle patterns of dispensability were evident within patients diagnosed with multiple phenotypically distinct cancers. For example, in a germline MSH6 carrier, two primary and clonally unrelated tumors arose: a colon cancer and an upper tract urothelial bladder cancer (both LS-associated cancer types). Tumor sequencing and confirmatory IHC indicated that the colorectal cancer retained WT MSH6 and was microsatellite stable (MSS), and thereby probably sporadic in origin. In contrast, the urothelial tumor acquired biallelic MSH6 inactivation via a second somatic truncating mutation, which drove MSI and somatic hypermutation (Fig. 5c). This finding reaffirms the obligate nature of somatic biallelic inactivation of the germline MMR allele for the MSI phenotype both across and within patients, even in LS-associated tumor types, with implications for therapy because MSI is a tumor-agnostic biomarker of response to immune checkpoint blockade.12

Overall, our analyses reveal that 27% of all cancers diagnosed in patients with high-penetrance pathogenic variants, similar to most cancers in carriers for lower-penetrance variants, were unrelated cancer types, retained heterozygosity and appeared to lack the somatic phenotypes indicative of dependence on the germline allele. Collectively, these analyses demonstrate how integration of concomitant germline and somatic tumor profiling can assist in distinguishing those tumors in which the pathogenic germline allele contributed to tumor formation or progression from those in which it probably represents an incidental finding (Fig. 6).

Discussion

A primary goal of cancer genomics has been the discovery of somatic drivers of tumorigenesis, and decades of investigation have revealed a broad spectrum of such changes of both biological and therapeutic importance. In parallel, germline investigation has largely focused on epidemiological and association-based studies that have lacked integrated somatic mutational profiles. In the present study, extending the dichotomy of driver versus passenger events in somatic tumorigenesis to germline cancer genetics, we have established at least two distinct routes to tumorigenesis in patients with germline pathogenic alleles (Fig. 6a).

In carriers of high-penetrance pathogenic variants, lineage-dependent selective pressure for biallelic inactivation was associated with earlier age of onset and specific somatic phenotypes indicative of dependence on the germline allele for tumorigenesis. In such patients, this germline ‘driver’ was likely the founding event that directly promoted cellular transformation and tumor initiation, ultimately shaping the somatic mutational profile of the resulting tumors, with subsequent somatic driver events arising to accelerate tumor formation, progression and potentially therapeutic sensitivity and resistance. On the other hand, 27% of all tumors diagnosed in patients with high-penetrance pathogenic variants, and most cancers arising in carriers associated with lower penetrance, neither represented associated lineages nor exhibited somatic loss of the WT allele, suggesting that the pathogenic germline variant did not contribute to the pathogenesis of most of these tumors. Even tumors that did exhibit biallelic inactivation often lacked somatic hallmarks of germline dependence when occurring in nonassociative cancer types, and tumors with pathogenic germline mutations in well-known cancer predisposition tumor-suppressor genes, such as BRIP1, NBN and MUTYH, exhibited rates of biallelic loss indistinguishable from those in carriers of benign variants, even in cancer types where these genes were shown to confer elevated risk. As these germline variants have been implicated by population-based studies, we cannot exclude the possibility that they may facilitate an environment that promotes tumor initiation which does not depend on loss of the WT allele. The resulting tumors nevertheless appear to be phenotypically indistinguishable from sporadic nonhereditary cancers, empirically extending decades-old models in the field.10,11

Overall, our study cohort presents a unique opportunity not only to gain a deeper understanding of, but also to re-evaluate prior findings in contexts that integrate somatic features.

Although previous studies have explored somatic phenotypes in carriers of germline susceptibility alleles for individual genes or cancer types, our study is designed to systematically characterize the role of germline pathogenicity in tumorigenesis across genes and lineages. A similar analysis was recently published for the pan-cancer TCGA cohort, demonstrating that tumor-specific zygosity alterations were prevalent for pathogenic germline variants and VUSs. However, our study overcomes key prior limitations of the TCGA dataset. First, ours represents a larger and more diverse cohort of matched tumor-normal sequencing data, enabling the study of rare, previously unexamined cancer types and biallelic inactivation of individual germline variants in susceptibility genes, which have historically been grouped together. Second, our dataset of germline variants incorporates and recapitulates clinical-grade pathogenicity assessment based on expert-guided criteria, ensuring the robustness and clinical applicability of our findings. Third, by classifying pathogenic variants by penetrance level, we were able to assess and identify the role of germline penetrance in influencing somatic phenotypes. Fourth, the high depth of coverage afforded by our FDA-authorized clinical sequencing assay (more than sixfold greater than conventional sequencing cohorts such as TCGA) enabled more precise and definitive zygosity estimates in matched tumors. Collectively, these key distinguishing features of our study allowed us to empirically recover a majority of established tumor-type associations with cancer susceptibility genes and systematically characterize the context-specific effects of zygosity, penetrance and lineage in shaping the role of germline variants in tumorigenesis.

These results have wide-ranging implications for the clinical management of cancers arising in carriers of pathogenic cancer-associated variants. Ultimately, they signal the need for a...
fundamental shift in current approaches to clinical assessment, whereby an integrated analysis of somatic and germline alterations is required to present a more complete view of a patient’s cancer (Fig. 6b). In this framework, somatic features such as biallelic inactivation and co-occurring mutational signatures in the arising tumor complement population frequency and family history to directly inform the interpretation of germline variants. Moreover, our population-scale dataset of germline variants and their zygosity in the corresponding tumors will provide important context to aid the prospective interpretation of VUSs in future patients and catalyze broader functional characterization of variants and cancer predisposition genes. To facilitate this change, we have established a comprehensive and searchable online portal of germline variants and their corresponding tumor-specific zygosity across common and rare cancer types from our study as a community resource (https://www.signaldb.org).

Our results also carry important implications for the treatment of patients harboring pathogenic germline alterations. From a therapeutic perspective, the recent emergence of immune checkpoint blockade and poly(ADP-ribose) polymerase (PARP) inhibitor therapy has led some to suggest broader germline testing as a guide to therapy selection. Whereas an MSI/MMR phenotype is established as a tissue-agnostic biomarker of immunotherapy response7,8, currently, tumor biopsy is the gold standard of tumor mutation burden assessment. As ground-truth functional data are lacking, particularly outside of a small number of well-characterized cancer susceptibility genes, our ability to annotate and predict pathogenic germline variants is constrained by the current state of biological knowledge. Furthermore, the systematic framework adopted in the study for pathogenicity assessment and penetration assignments does not capture the multitude of contexts in which these could be reinterpreted for a small number of variants. For example, germline mutations in SDHD, SDHAF2 and MAX have a parent-of-origin effect with disease prevalence associated in only the carriers with paternally inherited alleles, whereas such allele-specific lines of inheritance could not be ascertained in our cohort19. Moreover, our study data are limited by the targeted nature of our clinical sequencing assay and its focus on DNA-based alterations, allowing for the possibility of occult functional inactivation via alternative genomic or epigenetic mechanisms. We believe further careful examination of carriers of heterozygous pathogenic germline alleles will ultimately be necessary with additional orthogonal molecular characterization. Nevertheless, there is little evidence from prior studies of biallelic inactivation via monoallelic epigenetic silencing, which appears to be a rare event in germline carriers16. Although the weight of our data indicates that biallelic inactivation of genes harboring pathogenic germline alleles is generally a pivotal step to drive tumorigenesis and/or tumor progression, the extent to which haploinsufficiency plays a role in tumorigenesis of patients who harbor pathogenic variants in key genes, particularly those that maintain genome integrity, is only beginning to emerge20. Such a route to tumor initiation in preneoplastic cells, which have an endogenously increased rate of mutagenesis due to moderately impaired DNA-repair pathways, could be stochastically triggered by acquisition of one or more oncogenic somatic driver alterations. Furthermore, future studies could also elucidate the temporal contexts in which pathogenic variants are either essential or dispensable for seeding tumorigenesis. Finally, although we showed strong associations with age of onset among those carriers with high-penetration variants, as well as carriers among those who harbor biallelic inactivation, we cannot
exclude the role family history played in early diagnosis of tumors in these patients.

In conclusion, our results suggest that a new conceptual framework for understanding the heritable component of cancer is needed. Germline predisposition alleles among all penetrance categories are currently classified as pathogenic or likely pathogenic based on associations with elevated cancer incidence, yet we show in the present study that these variants may make little or no contribution to the pathogenesis and/or maintenance of the resulting cancers. The differential effects on cancer etiology for even high-penetrance genes by cancer type does not diminish the importance of cancer screening, genetic counseling or disease-specific preventive strategies, but emphasizes that not all cancers in patients with germline predisposition alterations have a biology related to that alteration. Although we chose to describe deleterious germline variants as 'pathogenic' to be consistent with existing conventions, these data suggest the need for a conceptual shift to better reflect the underlying biology of the resulting tumors which acknowledges lineage- and patient-specific variation in pathogenicity. Ultimately, careful integrated germline and somatic molecular characterization will be necessary to understand the role germline pathogenicity plays in the biology and therapeutic management of a given cancer.

Online content
Any methods, additional references, Nature Research reporting summaries, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00949-1.

Received: 16 November 2020; Accepted: 9 September 2021; Published online: 5 November 2021

References
1. Garber, J. E. & Offit, K. Hereditary cancer predisposition syndromes. J. Clin. Oncol. 23, 276–292 (2005).
2. Rahman, N. Realizing the promise of cancer predisposition genes. Nature 505, 302–308 (2014).
3. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. Proc. Natl Acad. Sci. USA 68, 820–823 (1971).
4. Nicholls, K. E., Malkin, D., Garber, J. E.,Fraumeni, J. F. & Li, F. P. Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. Cancer Epidemiol. Biomark. Prev. 10, 83–87 (2001).
5. Chandralekharapua, S. C. et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science 276, 404–407 (1997).
6. Moore, K. et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. N. Engl. J. Med. 379, 2495–2505 (2018).
7. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. N. Engl. J. Med. 372, 2509–2520 (2015).
8. Fong, P. C. et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N. Engl. J. Med. 361, 123–134 (2009).
9. Stadler, Z. K., Schrader, K. A., Vijai, J., Robson, M. E. & Offit, K. Cancer genomics and inherited risk. J. Clin Oncol. 32, 687–698 (2014).
10. Tung, N. et al. Counselling framework for moderate-penetrance cancer-susceptibility mutations. Nat. Rev. Clin. Oncol. 13, 581–588 (2016).
11. Jasperson, K. W., Troupe, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. Gastroenterology 138, 2044–2058 (2010).
12. Hampel, H. & Peltomaki, P. Hereditary colorectal cancer: risk assessment and management. Clin. Genet. 58, 89–97 (2000).
13. Couch, F. J. et al. Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncol. 3, 1190–1196 (2017).
14. Thavaneswaran, S. et al. Therapeutic implications of germline genetic findings in cancer. Nat. Rev. Clin. Oncol. 16, 386–396 (2019).
15. Zhang, J. et al. Germline mutations in predisposition genes in pediatric cancer. N. Engl. J. Med. 373, 2336–2346 (2015).
16. Huang, K.-L. et al. Pathogenic germ-line variants in 10,389 adult cancers. Cell 173, 355–370.e14 (2018).
17. Lu, C. et al. Patterns and functional implications of rare germ-line variants across 12 cancer types. Nat. Commun. 6, 10086 (2015).
18. Mandelker, D. et al. Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. JAMA 318, 825–835 (2017).
19. Jonsson, P. et al. Tumour lineage shapes BRCA-mediated phenotypes. Nature 571, 576–579 (2019).
20. Cheng, D. T. et al. Comprehensive detection of germline variants by MSK-IMPACT, a clinical diagnostic platform for solid tumor molecular oncology and concurrent cancer predisposition testing. BMC Med. Genom. 10, 33 (2017).
21. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–424 (2015).
22. Gryfe, R., Di Nicola, N., Gallinger, S. & Redston, M. Somatic instability of the APC I1307K allele in colorectal neoplasia. Cancer Res. 58, 4040–4043 (1998).
23. Win, A. K., Hopper, J. L. & Jenkins, M. A. Association between monoallelic MUTYH mutation and colorectal cancer risk: a meta-regression analysis. Fam. Cancer 10, 1–9 (2011).
24. Knudson, A. G. Two genetic hits (more or less) to cancer. Nat. Rev. Cancer 1, 157–162 (2001).
25. Nickerson, M. L. et al. Mutations in a novel gene lead to kidney tumors, lung adenocarcinomas, and breast cancer. Science 505, 302–308 (2014).
26. Baysal, B. E. et al. Mutations in the BRCA2 gene and breast cancer. Science 276, 848–851 (2000).
27. Polak, P. et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat. Genet. 49, 1476–1486 (2017).
28. Piñati, C. et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J. Pathol. 242, 10–15 (2017).
29. Viel, A. et al. A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. eBioMedicine 20, 39–49 (2017).
30. Scarpà, A. et al. Whole-genome landscape of pancreatic neuroendocrine tumours. Nature 543, 65–71 (2017).
31. Chakravarty, D. et al. Oncokb: a precision oncology knowledge base. JCO Precis. Oncol. https://doi.org/10.1200/POT.17.00011 (2017).
32. Le, D. T. et al.Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 357, 409–413 (2017).
33. Lynch, H. T., Shaw, M. W., Magnuson, C. W., Larsen, A. L. & Krush, A. J. Hereditary factors in cancer. Study of two large midwestern kindreds. Arch. Intern. Med. 117, 206–212 (1966).
34. Lynch, H. T. & Krush, A. J. Cancer family ‘G’ revisited: 1895–1970. Cancer 27, 1505–1511 (1971).
35. Latham, A. et al. Microsatellite instability is associated with the presence of Lynch syndrome pan-cancer. J. Clin. Oncol. 37, 286–295 (2019).
36. Möller, P. et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. Gut 67, 1306–1316 (2018).
37. Kinzler, K. W. & Vogelstein, B. Landscaping the cancer terrain. Science 280, 1036–1037 (1998).
38. Kinzler, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature 386, 761–763 (1997).
39. Baysal, B. E. et al. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. Science 287, 848–851 (2000).
40. Karayiayvaz-Yildirim, M. et al. Aneuploidy and a deregulated DNA damage response suggest haploinsufficiency in breast tissues of BRCA2 mutation carriers. Sci. Adv. 6, eaay2611 (2020).

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

Study cohort and prospective sequencing. The study cohort comprised 18,392 samples from 17,152 patients. All patients underwent prospective clinical sequencing as part of their clinical care (February 2014 to July 2017). The present study was approved by the MSK Cancer Center’s Institutional Review Board (IRB) and all patients provided written informed consent for tumor sequencing and review of medical records for demographic, clinical and pathology information. Genomic sequencing was performed on tumor DNA extracted from formalin-fixed, paraffin-embedded tissue and normal DNA extracted from mononuclear cells from peripheral blood in all patients. Patient samples were sequenced in a CLIA-compliant laboratory during routine clinical testing, using one of three versions of the MSK-IMPACT-targeted sequencing panel, which included 341- and 410-gene versions and the current 468-gene FDA-authorized panel using methods and somatic analysis described previously. This cohort comprised predominantly adult solid cancers, whereas 2.2% of all patients were children (<18 years) and 7.5% were children <18 years excluded from age-dependent analyses. The age of patients at clinical sequencing ranged from 18 years to 96 years (median 59 years), and female patients comprised 54% of our study cohort. Tumors were obtained from the primary site in 58% of patients and a metastatic site in 42% of patients, with variable prior treatment status.

Germline variant discovery and annotation. We performed germline variant calling using a sequence analysis pipeline validated for clinical use in a CLIA-compliant laboratory performing clinical sequencing of patient tumors and matched normal blood specimens to guide clinical care. As the aforementioned three versions of the MSK-IMPACT assay utilized in the present study represented incrementally larger panels without removing genes, all patients possessed sequencing data for the 341 genes in the original assay version, whereas 14,752 patients were also sequenced for the incremental 69 genes between the 341- and 410-gene versions. Germline variant discovery was performed in these 410 genes; the 58 genes added to the final assay version (468-gene version) were excluded from analyses due to insufficient sample size (n = 5,765). All frequencies reported throughout the study were calculated using their gene-specific sample size. A total of 3,358 patients had consented for identification analysis of germline variants at the time of clinical data freeze (June 2017) via an IRB protocol (NT01775072). Genomic and limited clinical data, including existing cancer diagnosis at the time of genetic analysis, for all other patients analyzed in the present study (n = 13,794) were anonymized before germline variant discovery and subsequent analyses. All single-nucleotide variants and small insertions and deletions (indels) identified in any blood normal sample were annotated with myvariant.info (as of August 2017). Population frequencies were obtained from gnomAD (r2.0.2). Curated assessments for pathogenicity were obtained from ClinVar (as of September 2017).

Germline CNV discovery. DNA copy-number variants (CNVs) were determined in all normal blood samples with a clinically validated pipeline as previously described. Briefly, germline CNVs were identified by comparing GC-normalized sequence coverage of targeted regions in normal samples with a standard diploid normal genome selected among a sequenced pool of normal specimens. To retain high-confidence germline CNVs, the following criteria were utilized: only thosegenic and intragenic deletions having a fold-change (log-transformed ratio) of normalized coverage between the normal specimen and the control genomic range of 1.5 and falsediscovery rate (FDR) of less than 0.01 were considered. Single-exon deletions were excluded as presumptive false positives and multixonic intragenic deletions were required to span contiguous exons. PMS2 deletions spanning exons 3, 13 and 14 were excluded due to pseudogene content. For all recurrent deletions, each event was examined in the corresponding tumor specimen for support and excluded from consideration as true germline deletions if discordant between normal and tumor specimens. Overall, 172 high-confidence germline deletions were included as pathogenic events to which we ascribed penetrance levels in a manner identical to individual sequence variants by gene (below).

Penetrance stratification for pathogenic variants. We stratified pathogenic variants by their gene-level estimate of penetrance using the relative risk (RR) of individuals with the variant developing cancer compared with the general population. All genes with pathogenic variants were classified into one of four categories based on community standards: high penetrance (RR ≥ 5), moderate penetrance (RR = 2–5), low penetrance (RR < 2) and uncertain penetrance for genes with pathogenic variants that are either not well characterized or have not been previously associated with germline predisposition to cancer. Genes were assigned to a penetrance category based on the highest reported risk in any cancer type. RR estimates were compiled from a literature review, including the National Comprehensive Cancer Network Clinical Practice guidelines for Breast, Ovarian and Colorectal cancers (www.nccn.org), and reviewed by medical geneticists at our institution to ultimately ascribe penetrance levels to each gene. We also performed penetrance stratification at the variant level when prior evidence clearly demonstrated a different penetrance level of a given variant from what is currently attributed to the gene. The following five variants had their penetrance level reassigned: EPCAM copy-number deletions were assigned high penetrance; similarly, CHEK2 1157T, FH K477dup and VHL R200W variants were reclassified as uncertain penetrance as ascribed by multiple prior studies. Although recent data suggest that FH K477dup provides a better variant in renal cell carcinoma, we assigned an uncertain penetrance to this variant to remain consistent with ClinVar. Finally, despite APC being classified as a high-penetrance gene in colorectal cancer, the APC 11307K is an Ashkenazi Jewish founder mutation that is more common, arising in 255 patients in our cohort. Although APC 11307K confers an elevated cancer risk with an OR of 2.1 (95% confidence interval (CI) = 1.64–2.86), its effect size is far less than loss-of-function APC variants and was therefore reassigned as low penetrance.

Assessing association by prevalence. To assess the association between the presence of germline pathogenic alleles in a given gene and cancer type, we developed an ancestry-controlled, permutation-based statistical framework. For a given cancer type (as documented at the time of genetic analysis), we tested the observed frequency of pathogenic variants in a given gene against a background distribution of frequencies generated from all other tumor types, ensuring that the underlying population structure of the background distribution matched that of the tested cancer type. We randomly sampled the dataset consisting of all other tumor types 10,000 times to generate a distribution of background rates of germline pathogenic variants and compared them with the observed rate of germline pathogenic variants within the cancer type of interest, to calculate an empirical P value of enrichment per gene that is population structure aware. For genes with variants in multiple penetrance categories, we estimated the distribution of penetrance levels across each of the cancer types. Finally, we corrected for multiple hypothesis testing using the Benjamini–Hochberg method and reported only those associations with a q value <0.15 (corresponding to an FDR of 15%). Due to the composition of our cohort, and the limited information regarding prior diagnoses of multiple independent cancers in the same individual, we were only able to consider associations for germline pathogenic variants in certain genes and cancer types. We therefore performed a literature survey and rescued 24 such high-confidence associations comprising 47 pathogenic variants (Supplementary Table 5).

Zygosity inference, modeling and enrichment analysis. We inferred somatic zygosity for all germline benign and pathogenic variants using locus-specific and allelic-specific DNA copy-number inference, tumor purity and the observed variant allele frequency (VAF) in the tumor. Each germline variant was determined to be heterozygous, biallelic (loss of the WT allele) or to have lost the mutant allele using the following framework. To initially determine whether a given germline variant is in allelic imbalance in the corresponding tumor specimen, we determined whether its observed somatic VAF was consistent with the expected VAF given the locus-specific, allelic-specific DNA copy number, which was calculated as:

\[
\phi \times mcn + (1 - \phi) \times tcn
\]

where \(\phi\) is the tumor purity and tcn and mcn are the total and minor copy number at the locus spanning the variant. Germline variants were considered heterozygous if their observed VAF was either (1) consistent with the expected VAF (within its 95% binomial CI) given balanced heterozygosity (tcn and mcn of either 2 and 1 or 4 and 2 in diploid and genome doubled tumors, respectively), or (2) less than 0.5 of the lower bound of the 95% CI of the expected VAF corresponding to a tcn and mcn of 3 and 1, respectively, which was either single copy gain of the mutant or WT allele. Germline variants in allelic imbalance of any kind were those with an observed VAF that was either within or greater than the 95% CI of the expected VAF corresponding to a copy-number state other than balanced heterozygosity. For allelically imbalanced germline variants, loss of the WT was determined as those with an observed VAF within the 95% CI (or greater than the lower bound of the 95% CI) of the expected VAF corresponding to an mcn equal to 0 (observed VAF is concordant with the expected VAF when the lesser allele has a copy number of 0). Loss of the mutant allele was determined as the reverse of this latter scenario. The zygosities of germline variants were considered indeterminate if the results from zygosity analyses if the: (1) variant was homozygous in the germline; (2) read depth of coverage in the normal blood specimen was <50; and (3) FACETS-derived total and minor copy number were not evaluable at the corresponding locus. Pathogenic variants were also considered biallelic in the tumor if no LOH was present, but a second clonal somatic truncating mutation was observed, which represented 8% of all biallelic inactivations.

To assess the enrichment for biallelic inactivation (combination of LOH or second somatic mutation) targeting the WT allele for loss in the tumors of carriers of germline pathogenic variants, the rate of such changes was compared with a background distribution of similar biallelic inactivation spanning all nonpathogenic variants (both benign variants and VUS). Biallelic inactivation was determined in the corresponding tumor specimens per patient in a manner identical to that of pathogenic variants as described above. For these nonpathogenic variants, a second somatic mutation was often not phaseable with the germline variant and was therefore assumed to arise in trans and
lead to biallelic activation. Although this is likely to arise on either allele with approximately equal frequency, this model produces a conservative estimate of the background rate for enrichment analyses. We also evaluated background rates for all variants not annotated in OncoKB as ‘benign’ or ‘likely benign’ with Fisher’s exact test. Normalized expression in germline carriers from TCGA with previously defined LOH were compared with those with probable epigenetic silencing of the remaining WT allele using the Mann– Whitney U-test.

The comparison of somatic loss of germline pathogenic alleles in cancer types that were associated with the germline allele versus not associated with the germline allele was performed using a binomial test for significance. Correction for multiple hypothesis testing was performed using the Benjamini–Hochberg method. Among carriers of germline MMR mutations, the rates of MSI in tumors with and without biallelic activation of the germline allele were compared using Fisher’s exact test. The MSI phenotype in germline MMR carriers who were biallelic in their corresponding tumor was compared across different cancer types (Lynch associated, secondary Lynch versus non-Lynch) using the \( \chi^2 \) trend test.

All statistical tests were performed using the R or Scipy python package.

**Data availability**

Study results including cohort-wide prevalence and zygosity of germline and somatic mutations are available at https://www.singaldb.org and may be subject to a registration process and certain terms of use specified at https://www.singaldb.org/terms, including that the results may be used only for noncommercial research purposes without a license agreement with MSI Cancer Center.

Germline variants and tumor-specific zygosity estimates are available from the National Center for Biotechnology Information dbGaP archive at accession no. phs001858.v1.p1. In addition, the following publicly available data were used: annotations indicating statistically significant somatic mutations were derived from Hotspots (http://www.cancerhotspots.org); biological effects, prognostic information and treatment implications of specific cancer gene alterations were obtained from OncoKB as of June 2018 (http://www.oncokb.org); variant-level annotations aggregated from data resources for germline alterations were obtained from myvariant.info as of August 2017 (https://myvariant.info); population frequencies for observed germline alterations were derived from gnomAD r2.0.2 (https://gnomad.broadinstitute.org); and annotations regarding the deleterious nature of known germline variants associated with breast cancer risk.

**References**

41. Cheng, D. T. et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* 17, 251–264 (2015).

42. Won, H. H., Scott, S. N., Brannon, A. R., Shah, R. H. & Berger, M. F. Detecting somatic genetic alterations in tumor specimens by exom capture and massively parallel sequencing. *J. Vis. Exp.* 80, e50710 (2013).

43. Xin, J. et al. High-performance web services for querying gene and variant annotation. *Genome Biol.* 17, 91 (2016).

44. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443 (2020).

45. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44, D862–D868 (2016).

46. Nakagawa, H. et al. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. *Cancer Res.* 64, 4721–4727 (2004).

47. Hayward, R. B. et al. Extensive gene conversion at the PMS2 DNA mismatch repair locus. *Hum. Mutat.* 28, 424–430 (2007).

48. Kilpivaara, O., Allhupuro, P., Vahteristo, P., Aaltonen, L. A. & Nevanlinna, H. CHEK2 I157T associates with familial and sporadic colorectal cancer. *J. Med. Genet.* 43, e54 (2006).

49. Kilpivaara, O. et al. CHEK2 variant I157T may be associated with increased breast cancer risk. *Int. J. Cancer* 111, 543–547 (2004).

50. Apostolou, P. & Papasotiriou, I. Current perspectives on CHEK2 mutations in breast cancer. *Breast Cancer* 9, 331–335 (2017).

51. Zhang, L. et al. Fumarate hydratase FH c.1431_1433dupAAA (p.Lys477dup) variant is not associated with cancer including renal cell carcinoma. *Hum. Mutat.* 41, 103–109 (2020).

52. Alam, N. A. et al. Genetic and functional analyses of FH mutations in multiple cutaneous and uterine leiomyomatosis, hereditary leiomyomatosis and renal cancer, and fumarate hydratase deficiency. *Hum. Mol. Genet.* 12, 1241–1252 (2003).
53. Gossage, L., Cartwright, E., Eisen, T. & Bycroft, M. A detailed analysis of von Hippel Lindau (VHL) mutations in sporadic clear cell renal carcinoma (ccRCC), VHL syndrome, and Chuvash polycythaemia. *J. Clin. Oncol.* 28, e15024–e15024 (2010).

54. Liang, J. et al. APC polymorphisms and the risk of colorectal neoplasia: a HuGE review and meta-analysis. *Am. J. Epidemiol.* 177, 1169–1179 (2013).

55. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B Methodol.* 57, 289–300 (1995).

56. Bielski, C. M. et al. Widespread selection for oncogenic mutant allele imbalance in cancer. *Cancer Cell* 34, 852–862.e4 (2018).

57. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713 (2017).

58. Chang, M. T. et al. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* 8, 174–183 (2018).

59. Vogelstein, B. et al. Cancer genome landscapes. *Science* 339, 1546–1558 (2013).

**Acknowledgements**

We thank our patients and their families for participating in this study. We thank the members of the Molecular Diagnostics Service in the Department of Pathology, the Marie Josee and Henry R. Kravis Center for Molecular Oncology, the Robert and Kate Niehaus Center for Inherited Cancer Genomics, and the Berger and Taylor labs for discussions and support. This work was supported by National Institutes of Health awards (nos. P30 CA008748 and R01 CA227534 to M.F.B.), the Breast Cancer Research Foundation (to M.E.R. and K.O.), the Fund for Innovation in Cancer Informatics (to J.G.) and Cycle for Survival.

**Author contributions**

P.S., C.B., M.F.B. and B.S.T. designed the study. P.S., C.B., P.J., S.S.C., A.L.R., A.V.P., C.M.B., M.F.B. and B.S.T. designed and performed data analysis. C.F., A.S., G.J., M.P., J.H., N.S., R.C., J. Galle, A.Z., M.I., D.M.H., D.B.S. and M.F.B. assisted with prospective genomic and clinical data collection and sample annotation. P.S., C.B., Y.K., S.M., J.V., K.A.C., M.I.C., M.E.W., D.M., O.C.B., J.Z., M.E.R. and Z.K.S. assisted with germline variant pathogenicity and penetrance annotation. S.O.S., I.D., X.L., J. Gao and N.S. assisted in the development of the SignalDB portal. J.S. assisted with pathology. P.S., C.B., M.F.B. and B.S.T. wrote the manuscript with input from all authors. M.F.B. and B.S.T. contributed equally as senior authors.

**Competing interests**

K.A.C. reports advisory or consulting activities with AstraZeneca, MSD Ireland and GSK Ireland, and receives honoraria from Pfizer. M.I.C. has had an advisory role with Pfizer. D.M.H. received personal fees from Chugai Pharma, Boehringer Ingelheim, AstraZeneca, Pfizer, Bayer, Debiopharm Group and Genentech, and grants from AstraZeneca, Puma Biotechnology, Loxo Oncology (now owned by Eli Lilly) and Bayer. L.Z. received honoraria from Future Technology Research LLC, Roche Diagnostics Asia Pacific, BGi and Illumina, and has family members with leadership positions and ownership interests in Decipher Medicine. M.E.R. reports honoraria from Research to Practice, Intellisphere and Physicians Education Resource; consulting and advisory activities for AstraZeneca, Daiichi Sankyo, Epic Science, Merck and Pfizer (all uncompensated), and Change Healthcare; institutional research funding from AbbVie, AstraZeneca, Merck and Pfizer; and editorial services for AstraZeneca and Pfizer. D.B.S. has consulted with and received honoraria from Pfizer, Loxo/Lilly Oncology, Illumina, Vividion Therapeutics, Scorpion Therapeutics, Fore Biotherapeutics and BioBridge Pharma. Z.K.S. has an immediate family member who serves as a consultant in Ophthalmology for Alcon, Adverum, Gyroscope Therapeutics Ltd, Neurogene and RegeneronBio, outside the submitted work. M.F.B. reports receiving research funding from Illumina and Grail and advisory board activities for Roche. B.S.T. reports advisory board activities for Boehringer Ingelheim and Loxo Oncology at Lilly, and research support from Genentech. The remaining authors declare no competing interests.

**Additional information**

Extended data are available for this paper at [https://doi.org/10.1038/s41588-021-00949-1](https://doi.org/10.1038/s41588-021-00949-1).

Supplementary information The online version contains supplementary material available at [https://doi.org/10.1038/s41588-021-00949-1](https://doi.org/10.1038/s41588-021-00949-1).

Correspondence and requests for materials should be addressed to Michael F. Berger.

Peer review information *Nature Genetics* thanks Stephen J. Chanock, Clare Turnbull and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).
Extended Data Fig. 1 | Variant discovery and pathogenicity classification. a) Schematic of the workflow for the germline variant discovery pipeline. b) Contribution of most important features used in classification of pathogenicity. c) Evidence for pathogenicity in classifier-based pathogenicity calls.

Proportion of variant calls predicted as pathogenic (first four columns) or benign (fifth column) that exhibited orthogonal evidence of pathogenicity from ClinVar (April, 2020), by medical geneticist review, or as truncating mutations in tumor suppressor genes (excluding last exon or within 50 amino acids at the C-terminus end of the protein).
Extended Data Fig. 2 | Ancestry inference. a) The inference of ancestry from polymorphic SNPs with sufficient coverage in the MSK-IMPACT assay (see Supplementary Note). b) Breakdown of pathogenic variants by ancestry subpopulation: EUR, European; ASJ, Ashkenazi Jewish; AFR, African/African American; ASN, East/South Asian; OTH, Other.
Extended Data Fig. 3 | Mechanisms of biallelic inactivation. Different classes of somatic alterations leading to biallelic loss in carriers of pathogenic variants are shown by gene. LOH, Copy number loss of heterozygosity.
Extended Data Fig. 4 | Biallelic inactivation via epigenetic means in TCGA germline carriers. 

**a)** The number of carriers of one of 310 pathogenic germline mutations in 17 high penetrance genes for which biallelic inactivation was apparent by promoter methylation (dark blue; n = 10 total, 3.2% of such patients). Data suggests that heterozygous carriers do not acquire biallelic inactivation via epigenetic silencing of the remaining allele in appreciable numbers as an alternative mechanism to LOH. 

**b)** Germline mutations and promoter methylation combined for BRCA1 in ovarian cancers and MLH1 in colorectal cancers indicate that they arise mutually exclusively in affected cancers. All data were acquired from the PanCancerAtlas of The Cancer Genome Atlas project (see Supplementary Note).
Extended Data Fig. 5 | Tumor-specific zygosity inference by various classes of variants and specimen types. a) The rate of biallelic inactivation in the tumors for pathogenic variants as well as multiple classes of non-pathogenic variants including all variants of unknown significance (VUS, as in Fig. 1b), all common variants (MAF > 5%), and all variants annotated as benign in ClinVar. b) The rates of biallelic inactivation in primary and metastasis samples compared with those in benign variants in the corresponding specimen types (primary or metastasis). c) By penetrance level and specimen type (primary or metastasis), the rate of somatic biallelic inactivation of pathogenic variants in cancer types that are either associated or not with increased prevalence in carriers. In gray are benign germline variants in the same genes and cancer types. Points represent biallelic rates for high (n = 714), moderate (n = 354) and low/uncertain (n = 1,353) pathogenic variants. Error bars are 95% binomial CIs. d) As in Main Text Fig. 2c, the rate of somatic biallelic inactivation of germline pathogenic variants in high penetrance genes by association with increased prevalence. Not shown here are those genes with no association to any cancer type or those genes with fewer than five pathogenic variants. Shown are the fraction with somatic biallelic inactivation among carriers of pathogenic variants (red) and benign variants (gray) within the same gene. Error bars are 95% binomial CIs.
Extended Data Fig. 6 | Somatic mutations in APC I1307K carriers. APC I1307K, classified as low penetrance (see Supplementary Note), is a T > A polymorphism that creates a hypermutable tract of eight adenines that increases the propensity for polymerase slippage leading to an additional insertion of adenine, which generates a frameshift. In our cohort, this somatic APC I1307fs* frameshift mutation that results from the aforementioned polymerase slippage on the allele carrying the germline I1307K variant occurred only in colorectal cancers. Seven of these eight colorectal cancers harbored a second somatic mutation leading to biallelic inactivation, which reaffirmed the ‘three-hit’ model for this variant.
Extended Data Fig. 7 | Age of onset in germline carriers by association with cancer type. Age of cancer diagnosis is shown for carriers of high penetrance pathogenic germline variants stratified by association with cancer type and zygosity. Data shown for 10,076 germline WT patients, 330 carriers in associated lineages (276 with biallelic loss) and 176 carriers in non-associated lineages (67 with biallelic loss). Linear regression adjusting for cancer type, specimen type (primary vs. metastasis), genomic instability and sex. For the boxplots, the center red line is the median, the lower and the upper hinges represent the first and third quartiles for the ages of onset. The upper and lower whiskers extend up to 1.5 * IQR (interquartile range) above and below the upper and lower hinges, respectively.
Extended Data Fig. 8 | Mutational signatures associated with germline alleles. Spectrum of mutational signatures in germline carriers of pathogenic alleles in the indicated genes. A signature is considered present if 30% or greater of all somatic mutations are attributed to it. Only signatures that are detected in at least 5% of the carriers are shown. In parentheses is the number of carriers of pathogenic alleles that had 10 or more somatic mutations for robust mutational signature inference.
Extended Data Fig. 9 | Cancer type-specific differences in somatic alterations in carriers. A) The gene-specific pattern of somatic alteration differences in GISTs among carriers of high penetrance alleles that are biallelic in the corresponding tumors (left) versus the rest. B) As in panel (A) but for breast cancers harboring canonical PIK3CA or CCND1 alterations (black and gray are carriers of germline alleles that are somatic biallelic or not in the corresponding cancers, respectively). Related to main text Fig. 3.
Extended Data Fig. 10 | Somatic phenotypes of MSI-positive tumors in germline MMR carriers. a) The proportion of germline MMR carriers among patients who presented with Lynch-associated cancers by gene altered, grouped by zygosity and MSI phenotype. b) Tumor mutational burden (TMB) for MSI tumors in carriers of pathogenic germline variants in MLH1, MSH2, MSH6, and PMS2 indicating no significant difference between the mutational burden of MSI tumors in carriers of different MMR gene mutations. For the boxplots, the center line is the median, the lower and the upper hinges represent the first and third quartiles. The upper and lower whiskers extend up to 1.5 * IQR (interquartile range) above and below the upper and lower hinges, respectively. c) MSIsensor score as a function of the proportion of indels among somatic mutations is shown for MSS tumors (black) compared to those germline pathogenic MSH6 (light blue), PMS2 (dark blue), MLH1 (light green), and MSH2 (dark green) carriers indicate gene-specific differences in their somatic mutational phenotype. Specifically, MSH6-mutant patients had a lower intensity of the MSI phenotype as measured by MSIsensor ($P = 2.1 \times 10^{-3}$, Mann Whitney U test) and a lower proportion of somatic indels in affected tumors ($P = 3.7 \times 10^{-7}$, Mann Whitney U test). d) Three distinct classes of somatic mutations (insertions, deletions, and substitutions) in the affected tumors of germline carriers of the indicated MMR genes (same as in panel c) indicates different germline MMR dysfunctions drive mutation class-specific differences in the somatic MSI phenotype. For the boxplots, the center line is the median, the lower and the upper hinges represent the first and third quartiles. The upper and lower whiskers extend up to 1.5 * IQR (interquartile range) above and below the upper and lower hinges, respectively. e) Immunogenic burden, determined as the ratio of total number of mutation derived epitopes that are strong binders to the total number of non-synonymous mutations, is shown for MSI tumors harboring pathogenic germline variants in MMR genes along with tumors that are germline wild-type (Non-carrier, including non-MSI tumors) with TMB > 20. Tumors from carriers of germline mutations in MLH1/MSH2 had significantly higher immunogenic burden than those tumors that are carriers of germline mutations in MSH6/PMS2 ($P = 1.2 \times 10^{-5}$, after adjusting for tumor type and sample type). Compared with Non-carriers, carriers of germline mutations in MLH1 and MSH2 had significantly higher immunogenic burden ($P = 1.3 \times 10^{-5}$, $P = 8 \times 10^{-4}$, respectively, Wilcoxon test) while carriers with germline mutations MSH6 and PMS2 did not differ from Non-carriers. For the boxplots, the center line is the median, the lower and the upper hinges represent the first and third quartiles. The upper and lower whiskers extend up to 1.5 * IQR (interquartile range) above and below the upper and lower hinges, respectively.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  Confirmed

☐   The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐   A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐   The statistical test(s) used AND whether they are one- or two-sided

*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*

☐   A description of any covariates tested

☐   A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐   A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐   For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

*Give P values as exact values whenever suitable.*

☐   For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐   For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐   Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

*Our web collection on statistics for biologists contains articles on many of the points above.*

Software and code

Policy information about availability of computer code

Data collection

No code was used for data collection.

Data analysis

The following tools were used to analyze the data in this study:

FACETS v0.5.6 - https://github.com/mskcc/facets
Hotspots - http://www.cancerhotspots.org
OncoKB (As of June 2018) - https://www.oncokb.org
myvariant.info (as of August 2017) - https://myvariant.info
gnomAD r2.0.2 - https://gnomad.broadinstitute.org
Clnvar (as of September 2017) https://www.ncbi.nlm.nih.gov/clnvar/
Random Forests Sci-kit learn 0.19.1 https://scikit-learn.org/stable/
Mutational signatures in cancer - www.github.com/mskcc/mutation-signatures
MSIsensor v0.6 - https://github.com/ding-lab/msisensor/
NetMHCpan v4.1 http://www.cbs.dtu.dk/services/NetMHCpan/
POLYSOLVER https://software.broadinstitue.org/cancer/pga/polysolver
custom scripts: https://github.com/taylor-lab/somatic-germline

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Study results including cohort-wide prevalence and zygosity of germline and somatic mutations are available at https://www.signaldb.org/ and may be subject to a registration process and certain terms of use specified at https://www.signaldb.org/terms, including that the results may only be used for non-commercial research purposes without a license agreement with Memorial Sloan Kettering Cancer Center. Germine variants and tumor-specific zygosity estimates are available from the NCBi dbGaP archive at accession number phs001858.v1.p1. Additionally, the following publicly available data were used: annotations indicating statistically significant somatic mutations were derived from Hotspots (http://www.cancerhotspots.org); biological effects, prognostic information, and treatment implications of specific cancer gene alterations were obtained from OncoKB as of June 2018 (http://www.oncokb.org); variant-level annotations aggregated from data resources for germline alterations were obtained from myvariant.info as of August 2017 (https://myvariant.info); population frequencies for observed germline alterations were derived from gnomAD v2.0.2 (https://gnomad.broadinstitute.org); and annotations regarding the deleterious nature of known germline variants and associated phenotypes were downloaded from ClinVar as of September 2017 (https://www.ncbi.nlm.nih.gov/clinvar).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The study cohort consisted of all patients sequenced through our clinical sequencing program during the dates of the study. Power analyses and statistical tests were performed to indicate where results are statistically significant. |
| Data exclusions | No data were excluded |
| Replication | As this is a single-cohort study, no replication analysis was performed. Reproducibility was assessed through comparisons to prior literature and public databases. |
| Randomization | No randomization was performed |
| Blinding | No blinding of data was performed |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
| ☑ | Antibodies |
| ☑ | Eukaryotic cell lines |
| ☑ | Palaeontology and archaeology |
| ☑ | Animals and other organisms |
| ☑ | Human research participants |
| ☐ | Clinical data |
| ☐ | Dual use research of concern |

Methods

| n/a | Involved in the study |
| ☑ | ChiP-seq |
| ☑ | Flow cytometry |
| ☑ | MRI-based neuroimaging |

Human research participants

Policy information about studies involving human research participants

Population characteristics

The cohort was comprised of 17,152 patients with advanced cancer. 57.5% of the tumors sequenced were from the primary
### Clinical data

Policy information about [clinical studies](#).

All manuscripts should comply with the [ICMJE guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

| Clinical trial registration | NCT01775072 |
|----------------------------|-------------|
| Study protocol             | The protocol for prospective tumor sequencing of advanced cancer patients is available at clinicaltrials.gov, NCT01775072. Our study represents a retrospective analysis of the clinical sequence data generated through that study, which was approved as a retrospective research protocol by the MSK institutional review board. |
| Data collection            | The study cohort was comprised of 18,392 samples from 17,152 patients. All patients underwent prospective clinical sequencing as part of their clinical care (February 2014 to July 2017) in the CLIA-compliant Molecular Diagnostics Service at MSKCC. |
| Outcomes                   | No outcomes reported |

Note that full information on the approval of the study protocol must also be provided in the manuscript.