Title
The context-specific role of germline pathogenicity in tumorigenesis

Journal
Nature Genetics, 53(11)

ISSN
1061-4036

Authors
Srinivasan, Preethi
Bandlamudi, Chaitanya
Jonsson, Philip
et al.

Publication Date
2021-11-01

DOI
10.1038/s41588-021-00949-1

Peer reviewed

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2022 May 05.

Published in final edited form as:

Nat Genet. 2021 November ; 53(11): 1577–1585. doi:10.1038/s41588-021-00949-1.

The context-specific role of germline pathogenicity in tumorigenesis

Preethi Srinivasan[1,8,9], **Chaitanya Bandlamudi**[2,8], **Philip Jonsson**[2], **Yelena Kemel**[3], **Shweta S. Chavan**[2], **Allison L. Richards**[2], **Alexander V. Penson**[4,5], **Craig M. Bielski**[4,5], **Christopher Fong**[2,5], **Aijazuddin Syed**[1], **Gowtham Jayakumaran**[1], **Meera Prasad**[1], **Jason Hwee**[1], **Selcuk Onur Sumer**[2], **Ino de Bruijn**[2], **Xiang Li**[2], **JianJiong Gao**[2], **Nikolaus Schultz**[2,4,5], **Roy Cambria**[7], **Jesse Galle**[7], **Semanti Mukherjee**[3,6], **Joseph Vijai**[3,6], **Karen A. Cadoo**[3,6], **Maria I. Carlo**[3,6], **Michael F. Walsh**[3,6], **Diana Mandelker**[1], **Ozge Ceyhan-Birsoy**[1], **Jinru Shia**[1], **Ahmet Zehir**[1], **Marc Ladanyi**[1,4], **David M. Hyman**[6,11], **Liying Zhang**[1,10], **Kenneth Offit**[3,6], **Mark E. Robson**[3,6], **David B. Solit**[2,4,6], **Zsofia K. Stadler**[3,6], **Michael F. Berger**[1,2,4,*], **Barry S. Taylor**[2,4,5,11]

[1]Department of Pathology, New York NY.

[2]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, New York NY.

[3]Robert and Kate Niehaus Center for Inherited Cancer Genomics, New York NY.

[4]Human Oncology and Pathogenesis Program, New York NY.

[5]Department of Epidemiology and Biostatistics, New York NY.

[6]Department of Medicine, New York NY.

[7]Research and Technology Management, Memorial Sloan Kettering Cancer Center, New York NY.

*Correspondence to: Michael F. Berger, PhD, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY, 10065, 646-888-3386, bergerm1@mskcc.org.

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.L., 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.F.W., D.M., O.C.B., L.Z., K.O., 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 Astra Zeneca, MSD Ireland, GSK Ireland, and 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 of Decipher Medicine. M.E.R. reports honoraria from Research to Practice, Intellisphere, and Physician's 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 Opththalmology for Alcon, Adverum, Gyroscope Therapeutics Limited, Neurogene, and RegenexBio, 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, Loxo Oncology at Lilly, and research support from Genentech. The remaining authors declare no competing interests.

[8]These authors contributed equally

[9]Present address: Stanford University School of Medicine, Palo Alto, CA.

[10]Present address: Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles (UCLA), Los Angeles, CA.

[11]Present address: Loxo Oncology, a wholly owned subsidiary of Eli Lilly, Inc.

Abstract

Human cancers arise from environmental, heritable, and somatic factors, but how these mechanisms interact in tumorigenesis is poorly understood. Studying 17,152 prospectively sequenced cancer patients, 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.

Inherited mutations in cancer susceptibility genes can predispose individuals to develop tumors[1,2]. Deleterious mutations in genes such as *TP53*, *RB1*, and *MEN1* underlie genetic syndromes exhibiting distinct spectra of cancer predisposition[3–5]. Moreover, mutations in a subset of cancer susceptibility genes including *BRCA1*, *BRCA2*, *MLH1*, and *MSH2* are also associated with clinical phenotypes that predict response to specific cancer therapies[6–8]. The recent validation of pathogenic germline variants as predictive biomarkers of drug response in an expanding number of tumor types has led to the rapid adoption of more expansive germline genetic testing to guide clinical management[9–13]. Despite this progress, the role that germline variants in cancer susceptibility genes play in the oncogenesis of non-syndromic cancers and their value as tumor agnostic predictive biomarkers of drug response have been challenging to ascertain[14]. Although large-scale population studies of germline variants have been the primary tool for defining disease associations and risk, these approaches do not simultaneously evaluate somatic alterations or mutational signatures and are therefore limited in their ability to establish the biological role of germline alterations in the tumorigenesis of cancers that arise in individual germline carriers.

Understanding the role of germline pathogenicity in tumorigenesis has become especially urgent with the advent of broad-based clinical genomic profiling in unselected patient cohorts[15–18]. These studies have identified a greater proportion of patients with pathogenic germline alterations in cancer susceptibility genes than previously anticipated. Yet, the extent to which germline variants contribute to the formation of tumors in these patients varies widely, and some tumors may represent sporadic cancers arising independently via other biological mechanisms. We reasoned that a systematic integrated analysis of

germline and somatic mutations in individuals with cancer could illuminate the role of germline pathogenicity on the biological, phenotypic, and ultimately therapeutic dependence of individual tumors.

To explore how germline and somatic alterations cooperate to promote cancer initiation and progression, we leveraged a unique, pan-cancer clinical cohort of matched tumor and normal genomic data. The deep sequence coverage from targeted clinical sequencing enables both the identification of pathogenic germline variants and the high-precision inference of their zygosity in the concomitant tumors that arise. Using these data, we previously demonstrated that tumor suppressor genes *BRCA1* and *BRCA2* exhibit lineage-specific variable rates of biallelic inactivation and associated somatic phenotypes in carriers of germline variants[19]. Here, we extend these findings to all cancer susceptibility genes to systematically characterize the role of pathogenic germline alleles on tumorigenesis, and we describe a novel 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 advanced cancer patients diagnosed with one of 55 broad cancer types and 413 histological subtypes (Supplementary Table 1). Germline blood and matched tumor tissue DNA were sequenced using 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 (see Methods)[20]. Germline variant calling was performed based on a sequence analysis pipeline validated for clinical use in a Clinical Laboratory Improvement Admendments (CLIA)-compliant laboratory[20]. 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±8% and 89±7%, respectively) on a cohort of 6,009 cancer patients whose germline data were prospectively curated by expert clinical molecular geneticists using American College of Medical Genetics (ACMG) guidelines for clinical interpretation[21] (Extended Data Fig. 1, see 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[10,18] and manual review (see 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, 4). The overall frequency of germline pathogenicity reflected, in part, the composition of ancestry in our patient cohort, which we estimated was comprised of 18% of

patients of Ashkenazi Jewish ancestry, in whom a higher frequency of founder pathogenic germline variants has been reported (Extended Data Fig. 2, Supplementary Note). Overall, we identified a pathogenic germline variant in genes conferring significantly elevated risk for cancer incidence (high and moderate penetrance) in 7.8% of patients (Supplementary Tables 3–4). 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[22]. 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[23]. A higher proportion of patients of Ashkenazi ancestry were carriers of high/moderate penetrance variants (11.4%) compared to 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 all known associations between cancer susceptibility genes and tumor types as well as emerging associations even in well studied genes such as BRCA2 (Fig. 1B, 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 disposition genes[24]. 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 five or more 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 (see 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 to a background distribution of tumors harboring germline variants in the same genes predicted to have no effect on fitness (i.e., benign variants and variants of unknown significance, see 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%; $P = 5.7\times10^{-76}$). 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 second-hit in *MSH2*, *NF1*, *APC* and *CDH1* (Extended

Data Fig. 3). Although we cannot exclude the possibility of additional cryptic non-coding alterations or promoter hypermethylation as further contributors to biallelic inactivation, multi-modality 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, $P = 2.8\times10^{-125}$) 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 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* [19]. We grouped cancer types according to whether or not they were associated with increased incidence of pathogenic germline variants in each gene, either from our own ancestry-adjusted association testing or prior population-based studies (see Methods, Supplementary Table 5). Here, the rate of biallelic inactivation in high penetrance genes for germline carriers with associated tumor types was 85% compared to only 43% among all other cancer types ($P = 2.6\times10^{-21}$; Fig. 2B). Although a similar pattern was evident among moderate penetrance genes (72 versus 34% in associated and non-associated cancer types, $P = 5\times10^{-5}$), no such relationship existed for low/uncertain penetrance genes, patterns that were true for both primary and metastatic tumors cohort-wide (Extended Data Fig. 5C).

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 to benign variants (Fig. 2C, Supplementary Table 6). However, not all canonical tumor suppressor genes reached statistical significance for pan-cancer biallelic inactivation, 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 only considering 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, 2/2 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 cancer[25]. Likewise, biallelic inactivation

was evident in 3/4 gastric cancers with germline *CDH1* mutations, which are associated with hereditary diffuse gastric cancer syndrome. However, in other genes such as *NBN* and *BRIP1*, the biallelic rates for pathogenic and benign variants were indistinguishable. This observation may relate in part to the absence of tumor types profiled in our study in which pathogenic variants in these genes are functional but nevertheless highlights the importance of considering additional contexts for interpreting variant significance. Other enrichments for biallelic loss were variant- rather than gene-specific, such as for rare pathogenic alleles in *APC* associated with high penetrance but not for the more common low penetrance I1307K variant (Extended Data Fig. 6). Similarly, in colorectal cancers, the rate of biallelic loss among carriers of germline mutations in *MUTYH* is identical to those with benign variants. Overall, our results indicate that different classes of pathogenic germline alleles have very different roles in the pathogenesis of cancers diagnosed in carriers and that these roles further vary among cancer types.

Phenotypic consequences of biallelic inactivation

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 penetrance genes with biallelic inactivation had a significantly younger age of disease onset compared to germline WT cancer patients ($P = 6\times10^{-14}$, 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[26]. Here, *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)[26–30]. The mutational spectra of cancers in high penetrance germline carriers therefore 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 phenotypes[19].

Examining the acquisition of additional somatic drivers, we explored the evolution of cancers in germline carriers and compared them to sporadic tumors. We found that the tumors in germline carriers of high penetrance pathogenic variants that underwent somatic biallelic inactivation were significantly less likely to harbor a somatic gain-of-function driver alteration (see Methods)[31] compared to either tumors that retained heterozygosity or tumors harboring a lower penetrance or no germline allele ($P = 1.16\times10^{-49}$, χ^2 trend

test; $P = 1.23\times10^{-10}$multivariate test adjusted for tumor type and sample type; Fig. 3B). This difference remained significant even after correcting for age and tumor stage. Among individual cancer types, this depletion of oncogenic drivers was most significant in breast cancer and gastrointestinal stromal tumors (Fig. 3C), where germline carriers had tumors of established subtypes distinct from those driven by the pathognomonic somatic oncogenic mutations that typify these cancers (Extended Data Fig. 9). This dichotomy was nevertheless lineage-specific, as pancreatic and colorectal cancers driven by pathogenic germline alleles harbored 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 penetrance genes ($P = 8.3\times10^{-7}$, 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 non-associated high penetrance genes (16/23) ($P = 0.019$) and tumors in germline WT patients (891/1480) ($P = 0.025$).

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 penetrance genes that undergo biallelic inactivation. These cancers require fewer somatic oncogenic dependencies to confer a selective growth advantage, emphasizing the distinct etiologic role of the germline allele in disease pathogenesis. The pathogenic germline variant is thus likely 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 blockade[32]. In 118 patients (0.68%) who harbored a germline MMR allele diagnostic for LS[33–35], we characterized the MMR gene status in the corresponding tumor via zygosity inference and/or immunohistochemistry (IHC) staining for MMR protein expression. Altogether, somatic biallelic inactivation of the associated MMR gene was near-obligate for the MSI phenotype, with 80% of biallelic tumors and only 3% of heterozygous tumors exhibiting MSI (odds ratio = 124, $P = 9.4\times10^{-15}$; 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 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 types[35], 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 ($P = 2.5\times10^{-12}$).

Considering each MMR gene separately, the rate of biallelic inactivation varied widely ($P = 3.1\times10^{-7}$, Fig. 4C), ranging from 95% for *MSH2* to 33% for *PMS2*. Although the overall proportion of MSI tumors also varied by gene altered ($P = 4.9\times10^{-6}$), 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 ($P = 4.6\times10^{-6}$, Fig. 4D). Carriers of LS mutations in *MLH1* or *MSH2* were far more likely to have a canonical Lynch-associated tumor as compared to carriers of *MSH6* or *PMS2* alterations ($OR = 5.68$, $P = 5.3\times10^{-5}$), consistent with prior reports highlighting the lower lifetime risk of cancer in LS patients with *MSH6* or *PMS2* mutations compared to *MLH1* or *MSH2* mutations[36]. Additionally, tumors of *MSH6* carriers in LS-associated cancers had a lower rate of MSI (67%) compared to *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 TMB rates comparable to the other MMR genes, we observed a markedly lower intensity of the MSI phenotype as measured by MSIsensor ($P = 7\times10^{-3}$) and a lower proportion of somatic indels in the affected tumors ($P = 6\times10^{-5}$, Extended Data Fig. 10B–D). We also observed a lower immunogenic burden in *MSH6* tumors compared to *MLH1/MSH2* carriers as measured by the number of HLA-binding neopeptides generated per non-synonymous mutation (see 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 that delete 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, though this was least common in patients who had germline pathogenic alterations in higher penetrance genes ($P = 1.6\times10^{-7}$, Cochran-Armitage test for trend; Fig. 5A). Moreover, the rate of somatic loss of the pathogenic germline allele was significantly greater in non-associated cancer types ($P = 1.5\times10^{-9}$, Fig. 5B). These data suggest that in some patients, their cancers either lose dependence on a pathogenic germline allele during tumor progression, or alternately, 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 immunohistochemistry

indicated that the colorectal cancer retained WT *MSH6* and was microsatellite stable, thereby likely 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 as MSI is a tumor-agnostic biomarker of response to immune checkpoint blockade[32].

Overall, our analyses reveal that 27% of all cancers diagnosed in patients with high penetrance pathogenic variants, like the majority of 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 likely 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 significance. In parallel, germline investigation has largely focused on epidemiological and association-based studies that have lacked integrated somatic mutational profiles. Here, 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 the majority of 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 a majority of these tumors. Even tumors that did exhibit biallelic inactivation often lacked somatic hallmarks of germline dependence when occurring in non-associated 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 that does not depend on loss of the

WT allele. The resulting tumors nevertheless appear to be phenotypically indistinguishable from sporadic non-hereditary cancers, empirically extending decades-old models in the field[37,38]. Overall, our study cohort presents a unique opportunity to not only gain a deeper understanding of but also to reevaluate 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[16], demonstrating that tumor-specific zygosity alterations were prevalent for pathogenic germline variants and variants of unknown significance. However, our study overcomes key prior limitations of the TCGA dataset. First, ours represents the largest and most diverse cohort of matched tumor-normal sequencing data analyzed to date, enabling the study of rare, previously unexamined cancer types and of 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 6-fold 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 in order 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 variants of uncertain significance 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 adenosine diphosphate-ribose polymerase (PARP)

inhibitor therapy has led some to suggest broader germline testing as a guide to therapy selection. Whereas a MSI/MMR phenotype is established as a tissue-agnostic biomarker of immunotherapy response[7], our data argue that the presence of a germline MMR-associated alteration is not sufficient to guide therapy. The variable impact on tumorigenesis for even high penetrance genes also suggests care in the tumor lineage and patient-specific use of PARP inhibitor therapy[19], as not all tumors that arise in carriers of pathogenic germline variants in effectors of homologous recombination will be dependent on these lesions and be predicted to respond to PARP inhibition. The presence of heterozygous germline variants associated with high penetrance that seemingly do not directly promote tumor formation or progression argues that variant-level annotation of pathogenicity alone is likely insufficient to predict the impact on tumorigenesis and therapeutic response in a given patient without concomitant tumor profiling.

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 penetrance 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 cohort[39]. 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 alternate 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 carriers[16]. Although the weight of our data indicate 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 emerge[40]. Such a route to tumor initiation in pre-neoplastic cells that have 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 penetrance variants as well as carriers among these that 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 across all penetrance categories are currently classified as pathogenic or likely pathogenic based on associations with elevated cancer incidence, yet we show here that these variants may

have 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 preventative 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 that 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.

Methods

Study cohort and prospective sequencing

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). This study was approved by the Memorial Sloan Kettering 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 (FFPE) 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 with 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 as described previously [41,42]. This cohort was comprised of predominantly adult solid cancers, while 2.2% of all patients were pediatric (<18 years of age). Pediatric patients were excluded from age-dependent analyses. Age of patients at clinical sequencing ranged from 18 to 96 years (median 59 years). 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 [20,41]. As the aforementioned three versions of the MSK-IMPACT assay utilized here represented incrementally larger panels without removing genes, all patients possessed sequencing data for the 341 genes in the original assay version, while 14,752 patients were also sequenced for the incremental 69 genes between the 341- and 410-gene assay 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 analysis 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 identified 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 here (n=13,794) were anonymized prior to 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) [43]. Population frequencies were obtained from gnomAD (r2.0.2) [44]. Curated assessments for pathogenicity were obtained from ClinVar (as of Sept. 2017) [45].

Germline copy number variant discovery

DNA copy number variants were determined in all normal blood samples with a clinically validated pipeline as previously described [20,41]. Briefly, germline copy number variants (CNVs) were identified by comparing GC-normalized sequence coverage of targeted regions in normal samples to 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 those genic and intragenic deletions having a fold change (log-transformed ratio of normalized coverage between the normal specimen and the control genome) of less than −1.5 and false discovery rate-corrected $P < 0.01$ were considered. Single-exon deletions were excluded as presumptive false positives and multi-exonic intragenic deletions were required to span contiguous exons. *PMS2* deletions spanning exons 3, 13, and 14 were excluded due to pseudogene content [46,47]. 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 (see 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 to the general population. All genes with pathogenic variants were classified into one of four categories based on community standards [10,18]: 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. Relative risk estimates were compiled from literature review including the National Comprehensive Cancer Network (NCCN) Clinical Practice guidelines for Breast, Ovarian and Colorectal cancers (http://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 than 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* I157T, *FH* K477dup, and *VHL* R200W variants were reassigned as uncertain penetrance as ascribed by multiple prior studies[48–53]. Although recent data indicates that *FH* K477dup is likely a benign variant

in renal cell carcinoma, we have assigned an uncertain penetrance to this variant to remain consistent with ClinVar[45]. Finally, despite *APC* being classified as a high penetrance gene in colorectal cancer, the *APC* I1307K is an Ashkenazi Jewish founder mutation that is more common, arising in 255 patients in our cohort. Although *APC* I1307K confers an elevated cancer risk with an odds ratio of 2.1 (95% CI: 1.64 – 2.86) [54], 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 to 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 tested for association by gene and penetrance level in each of the cancer types. Finally, we corrected for multiple hypothesis testing using the Benjamini-Hochberg method [55] and reported only those associations with a q-value<0.15 (corresponding to a false discovery rate 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 underpowered to detect a subset of previously associations for prevalence of pathogenic variants in certain genes and cancer types. We therefore performed 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 allele-specific DNA copy number inference, tumor purity, and the observed variant allele frequency (VAF) in the tumor [19]. Each germline variant was determined to be either heterozygous, biallelic (loss of the WT allele), or having 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 if its observed somatic VAF was consistent with the expected VAF given the locus-specific allele-specific DNA copy number, which was calculated as:

$$\frac{\Phi * mcn + (1 - \Phi)}{\Phi * tcn + 2 * (1 - \Phi)}$$

where ϕ 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 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 whose observed VAF was either within or greater than the 95% CI of the expected VAF corresponding to a copy number state other than balanced heterozygosity [56]. For allelically imbalanced germline variants, loss of the WT was determined as those whose observed VAF was within the 95% CI (or greater than the lower bound of the 95% CI) of the expected VAF corresponding to an *mcn* equal to zero (observed VAF is concordant with the expected VAF when the lesser allele has a copy number of zero). Loss of the mutant allele was determined as the reverse of this latter scenario. The zygosity of the germline variant was considered indeterminate and excluded from zygosity analyses if the: 1) variant was homozygous in the germline; 2) read depth of coverage in the normal blood specimen was less than 50; 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 inactivation.

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 to a background distribution of similar biallelic inactivation spanning all non-pathogenic variants (both benign and variants of uncertain significance, 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 non-pathogenic 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 annotated in ClinVar as ‘benign’ with two or more gold stars, indicating that at least two submitters asserted non-pathogenic status. Finally, we assessed the background rate for only common variants (minor-allele frequency > 5%). The distribution of zygosity changes are identical between the three sets (Extended Data Fig. 5A).

Somatic mutational analyses and prioritization

Somatic mutations including substitutions and small insertions and deletions, gene-level focal DNA copy number amplifications and deletions (CNAs), and structural rearrangements were identified with a clinically validated pipeline as previously described [20,57]. All somatic alterations were further classified as oncogenic or likely oncogenic using OncoKB (June 2018), a curated knowledge base that provides evidence-based information about and classification of individual somatic mutations, CNAs, and structural alterations (http://oncokb.org/) [31]. Somatic mutations not classified by OncoKB as oncogenic, but have been previously identified as hotspots arising more frequently than expected in the absence of selection in population-scale cancer genome data, were similarly classified as likely oncogenic [58]. We also classified genes by function (oncogenes or tumor suppressors) based on OncoKB classification [31]. For the subset of cases in which biallelic inactivation was due

to a second somatic sequence variant (rather than LOH), only oncogenic or likely oncogenic mutations of all OncoKB levels or truncating mutations (frameshift insertions, deletions, and nonsense mutations) were considered.

For the purposes of analyzing the rate of somatic oncogenic alterations in germline carriers, we used only mutations, CNAs, and fusions that were classified as oncogenic (levels 1–4 per OncoKB). Here, we included 51 well-established oncogenes [59] that were present in all MSK-IMPACT assay versions. Tumors with somatic hypermutation (n=1,545) were defined as those with mutational burden above the 90th percentile of all tumors (which also included MSI tumors) and were excluded from this analysis. Germline copy number deletions were also excluded from this analysis due to the ambiguity in determining tumor-specific zygosity. To account for differences in the distribution of cancer types among the carriers of germline pathogenic variants in each penetrance class analyzed, we randomly sampled 100 times from the population of germline WT tumors to ensure we compared populations with a similar distribution of cancer types (referred to as WT type-matched). The mean and 95% binomial CIs are reported here (Fig. 3B). For assessing the number of somatic driver alterations per patient, we expanded to include all known/likely drivers [all oncogenic, likely oncogenic, and predicted oncogenic mutations, focal DNA copy number alterations, and fusions (levels 1–4 per OncoKB)] in any cancer gene, as well as any truncating mutation in a tumor suppressor gene [31]. The number of somatic driver alterations per patient for the germline WT patients was calculated using the same cohort of patients from which the random sampling was performed to ensure a similar distribution of tumor types to the high penetrance germline carriers whose tumors had biallelic inactivation. This ensured that the differences in the rate of somatic oncogenic drivers between the different groups were not driven primarily by differences in the cancer types among these patients.

Quantification and Statistical Analysis

Enrichment for biallelic inactivation (loss of WT) in Fig. 2 was assessed using logistic regression analyses. In Fig. 2A, for each penetrance category, we tested for enrichment for loss of WT with: $pathogenic \sim zygosity + cancer\_type + specimen\_type + genomic\_instability + sex$. $zygosity$ is encoded as ‘None’, ‘loss of Mutant’ and ‘loss of WT’. $specimen_{type}$ is ‘primary’ or ‘metastasis’. $genomic\_instability$ is calculated as the fraction of the genome that does not have the ‘normal’ inherited configuration of alleles (that is, a total copy number of 2 and minor copy number of 1). For tumors with whole genome doubling we also considered a total copy number of 4 and minor copy number of 2 as ‘normal’. Similarly, for Fig. 2B, we tested for enrichment of loss of WT in associated vs. not-associated for each of the penetrance categories with: $associated \sim zygosity + cancer\_type + specimen\_type + genomic\_instability + sex$. In Fig. 2C, we tested 57 tumor suppressor genes with five or more pathogenic variants for enrichment for loss of WT with: $pathogenic \sim zygosity + cancer\_type + specimen\_type + genomic\_instability$. Benjamini-Hochberg correction is applied to the p-values.

The number of somatic gain-of-function driver alterations (mutation, CNA, or fusion) [31] across tumors in different categories of germline carriers and non-carriers were compared using χ^2 trend test. Among individual tumor types, the comparison between somatic gain-of-function drivers between carriers of high penetrance germline alterations who undergo biallelic inactivation in tumors versus all other groups was performed using Fisher's exact test. The number of somatic drivers including gain-of-function and loss-of-function alterations were compared across different groups using ANOVA.

Differences in the age of cancer diagnosis between carriers and non-carriers of pathogenic germline alleles by penetrance, zygosity and sample type were assessed using log-rank test. The comparison of cases harboring germline alleles that got deleted in the incumbent clone of the resulting tumor by LOH targeting the mutant allele across different penetrance categories was performed using χ^2 trend test. The comparison of somatic loss of germline pathogenic alleles in cancer types that were associated with the germline allele vs. not associated with the germline allele was performed using Fisher's exact test. Normalized expression in germline carriers from TCGA with previously defined LOH [16,17] were compared with those with likely epigenetic silencing of the remaining WT allele using Mann Whitney test.

Enrichment of specific mutational signatures in carriers of germline pathogenic alleles in individual genes was performed using a binomial test for significance. Correction for multiple hypothesis testing was performed using Benjamini-Hochberg method [55]. 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 that were biallelic in their corresponding tumor was compared across different cancer types (Lynch-associated, secondary Lynch vs. Non-Lynch) using χ^2 trend test.

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

DATA AVAILABILITY

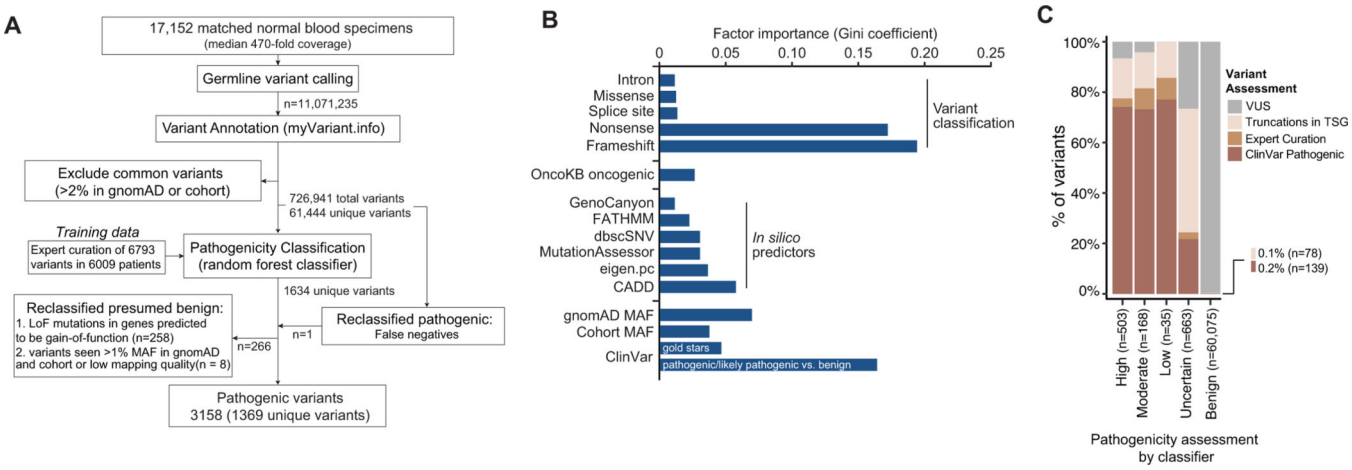
Study results including cohort-wide prevalence and zygosity of germline and somatic mutations are available at https://www.signaldb.org/ and may be subject to a registration process and certain terms of use specified at https://www.signaldb.org/terms, including that the results may only be used for non-commercial research purposes without a license agreement with Memorial Sloan Kettering Cancer Center. Germline 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 r2.0.2 (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).

CODE AVAILABILITY

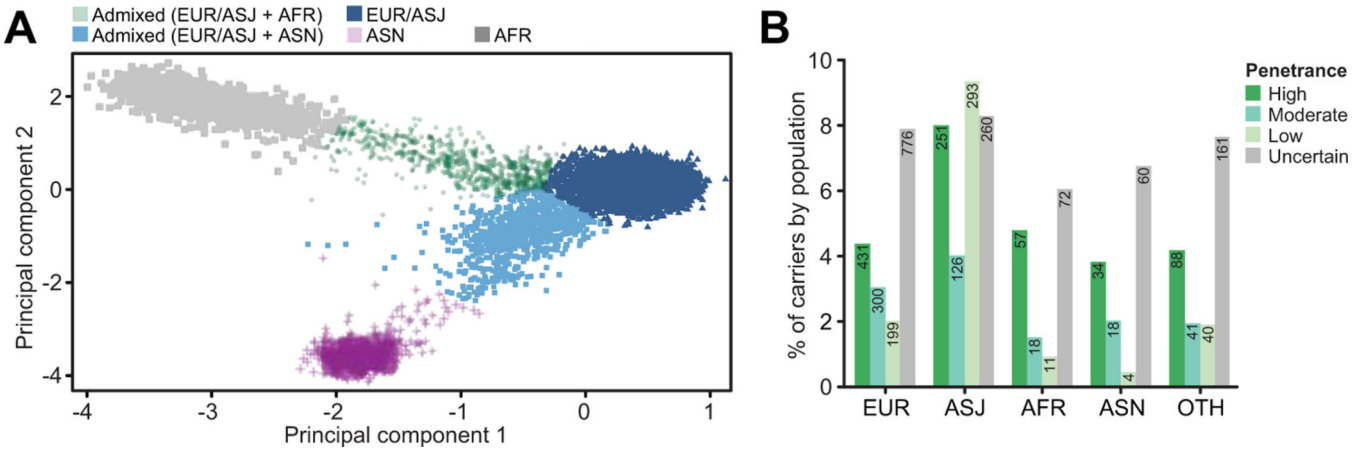
Source code is available at https://github.com/taylor-lab/somatic-germline.

Extended Data



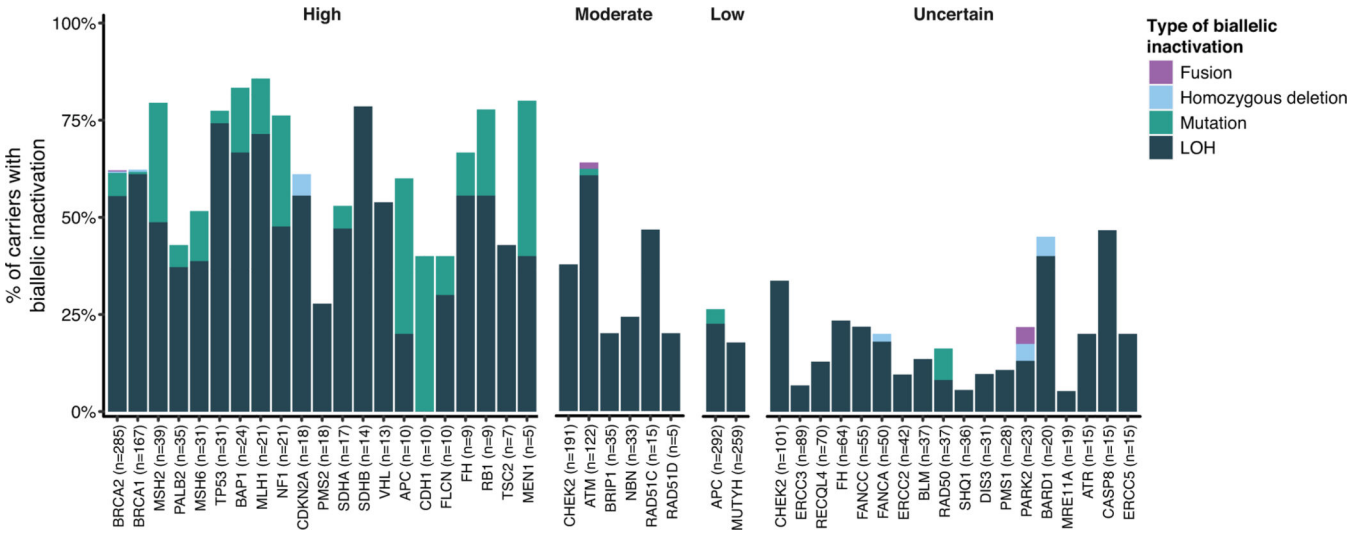
Extended Data Fig. 1: Variant discovery and pathogenicity classification.

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).

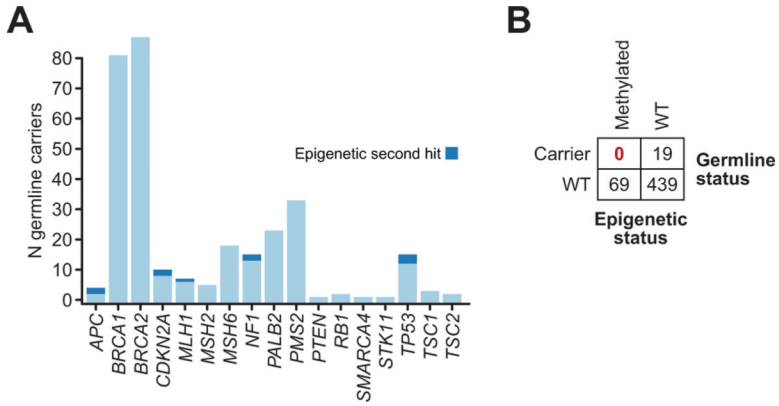


Extended Data Fig. 2: Ancestry inference.

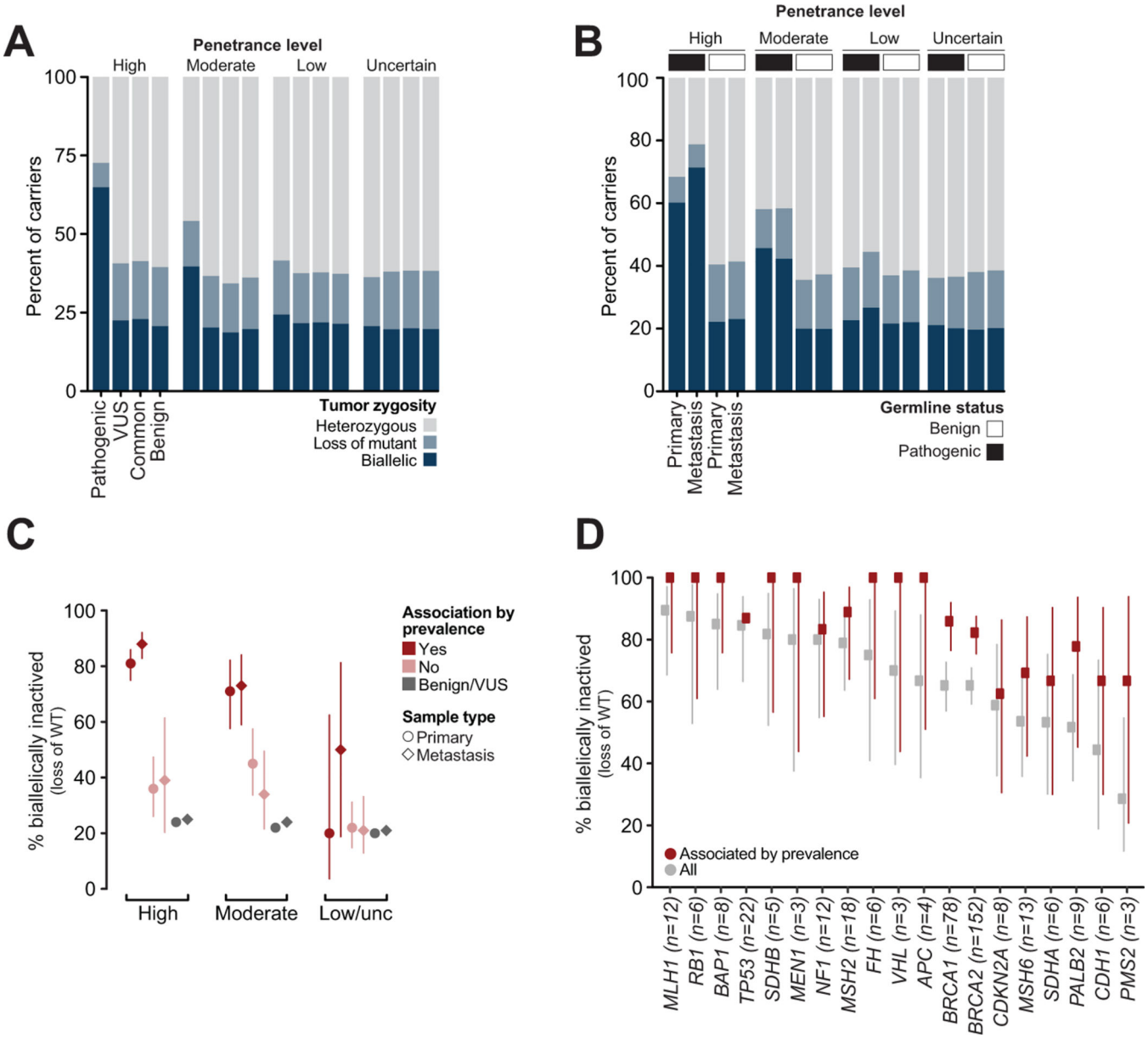
A) The inference of ancestry from polymorphic SNPs with sufficient coverage in the MSK-IMPACT assay (see Supplementary Note). **B)** Breakdown of pathogenic variants by ancestry subpopulation. European (EUR), Ashkenazi Jewish (ASJ), African/African American (AFR), East/South Asian (ASN), Other (OTH).



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. Copy number loss of heterozygosity (LOH).



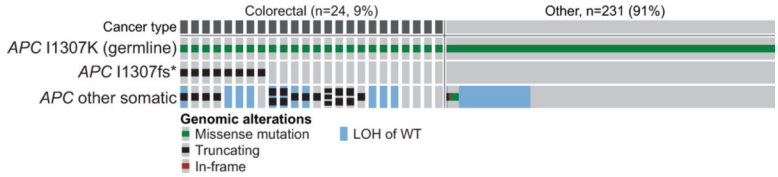
Extended Data Fig. 4: Biallelic inactivation via epigenetic means in TCGA germline carriers.
A) The number of carriers of one of 310 pathogenic germline mutations in 17 high penetrance genes for which biallelic inactivation was apparent by promoter methylation (dark blue; n=10 total, 3.2% of such patients). Data suggests that heterozygous carriers do not acquire biallelic inactivation via epigenetic silencing of the remaining allele in significant 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).



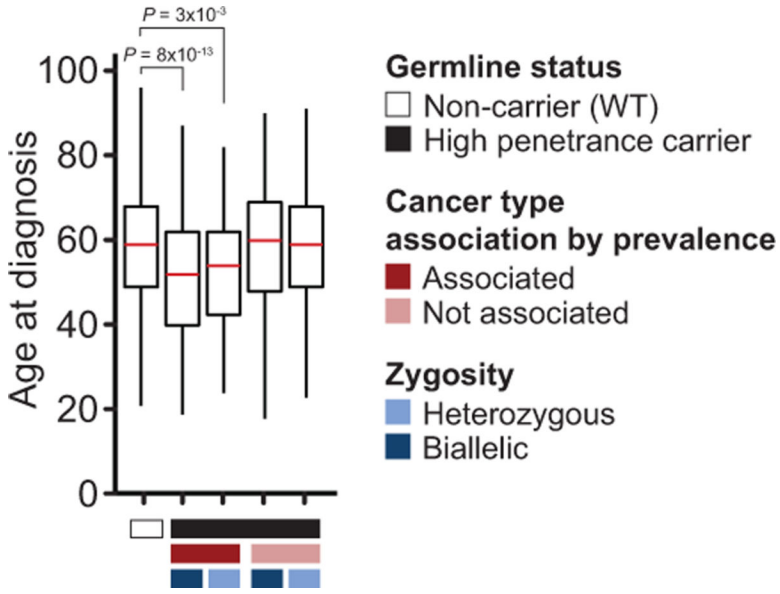
Extended Data Fig. 5: Tumor-specific zygosity inference by various classes of variants and specimen types.

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 confidence

intervals. **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 confidence intervals.

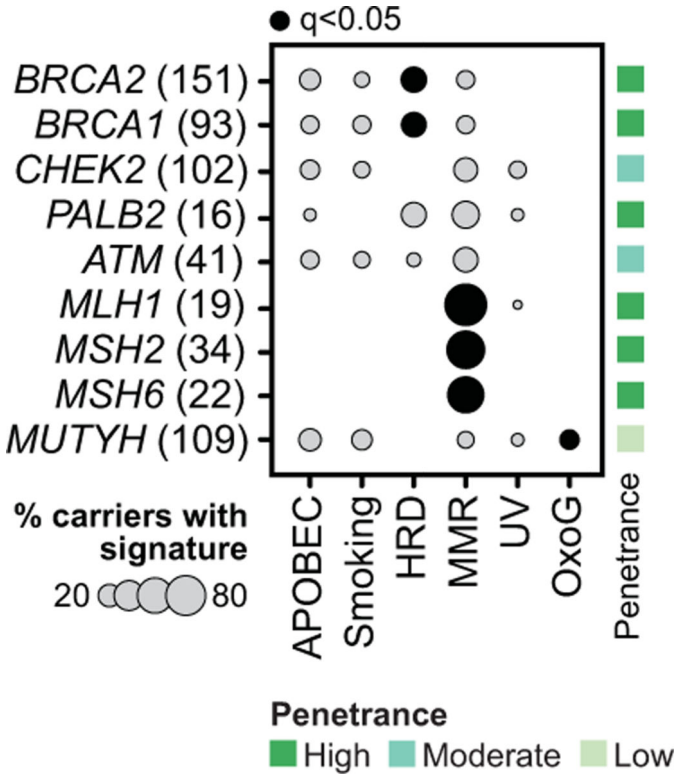


Extended Data Fig. 6: Somatic mutations in *APC* I1307K carriers.
APC I1307K, classified as low penetrance (see Supplementary Note), is a T>A polymorphism that creates a hypermutable tract of eight adenines that increases the propensity for polymerase slippage leading to an additional insertion of adenine, 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.

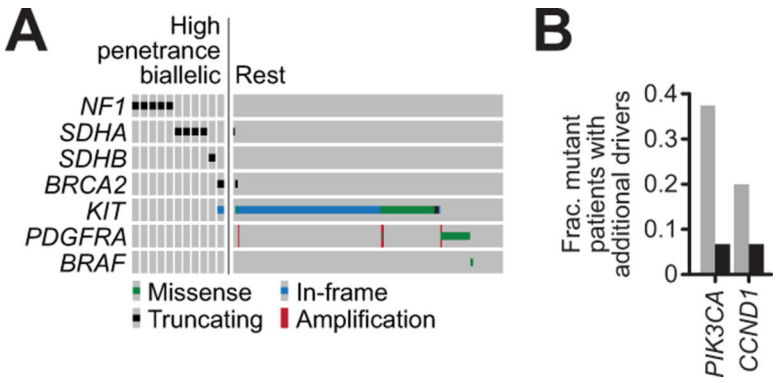


Extended Data Fig. 7: Age of onset in germline carriers by association with cancer type.
Age of cancer diagnosis is shown for carriers of high penetrance pathogenic germline variants stratified by association with cancer type and zygosity. Data shown for 10,076 germline WT patients, 330 carriers in associated 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.

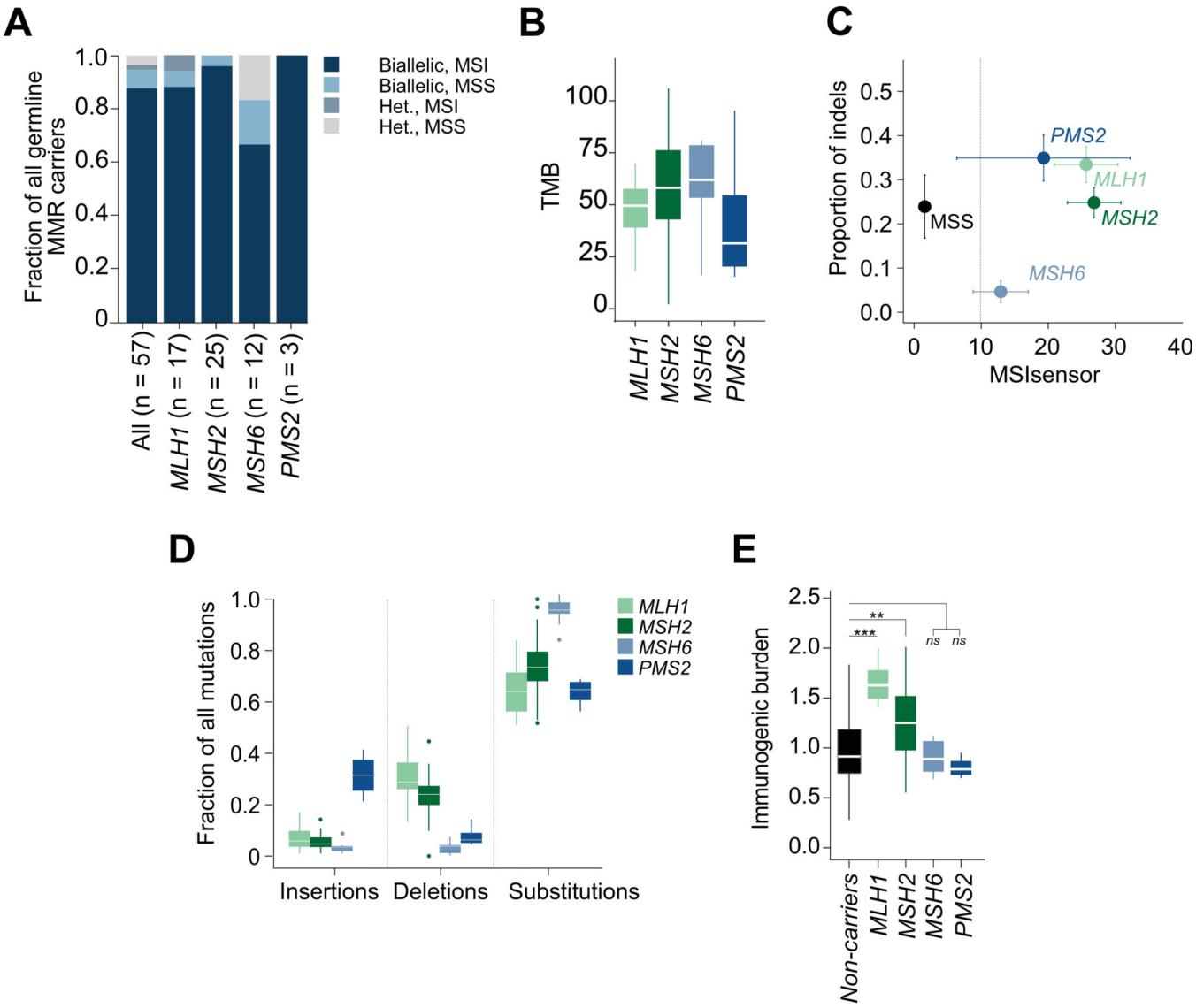


Extended Data Fig. 8: Mutational signatures associated with germline alleles.
Spectrum of mutational signatures in germline carriers of pathogenic alleles in the indicated genes. A signature is considered present if 30% or greater of all somatic mutations are attributed to it. Only signatures that are 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 Figure 3.



Extended Data Fig. 10: Somatic phenotypes of MSI-positive tumors in germline MMR carriers.

A) The proportion of germline MMR carriers among patients 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank our patients and their families for participating in this study. We gratefully acknowledge the members of the Molecular Diagnostics Service in the Department of Pathology, the Marie-Josée 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 P30 CA008748 and R01 CA227534 (M.F.B.), the Breast Cancer Research Foundation (M.E.R., K.O.), the Fund for Innovation in Cancer Informatics (J.G.), and Cycle for Survival.

REFERENCES

1. Garber JE & Offit K. Hereditary cancer predisposition syndromes. J. Clin. Oncol. 23, 276–292 (2005). [PubMed: 15637391]
2. Rahman N. Realizing the promise of cancer predisposition genes. Nature 505, 302–308 (2014). [PubMed: 24429628]
3. Knudson AG Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA 68, 820–823 (1971). [PubMed: 5279523]
4. Nichols KE, Malkin D, Garber JE, Fraumeni JF & Li FP Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. Cancer Epidemiol. Biomarkers Prev. 10, 83–87 (2001). [PubMed: 11219776]

5. Chandrasekharappa SC et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science 276, 404–407 (1997). [PubMed: 9103196]
6. Moore K. et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. N. Engl. J. Med. 379, 2495–2505 (2018). [PubMed: 30345884]
7. Le DT et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. 372, 2509–2520 (2015). [PubMed: 26028255]
8. Fong PC et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N. Engl. J. Med. 361, 123–134 (2009). [PubMed: 19553641]
9. Stadler ZK, Schrader KA, Vijai J, Robson ME & Offit K. Cancer genomics and inherited risk. J. Clin. Oncol. 32, 687–698 (2014). [PubMed: 24449244]
10. Tung N. et al. Counselling framework for moderate-penetrance cancer-susceptibility mutations. Nat. Rev. Clin. Oncol. 13, 581–588 (2016). [PubMed: 27296296]
11. Jasperson KW, Tuohy TM, Neklason DW & Burt RW Hereditary and familial colon cancer. Gastroenterology 138, 2044–2058 (2010). [PubMed: 20420945]
12. Hampel H. & Peltomaki P. Hereditary colorectal cancer: risk assessment and management. Clin. Genet. 58, 89–97 (2000). [PubMed: 11005140]
13. Couch FJ et al. Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncol. 3, 1190–1196 (2017). [PubMed: 28418444]
14. Thavaneswaran S. et al. Therapeutic implications of germline genetic findings in cancer. Nat. Rev. Clin. Oncol. 16, 386–396 (2019). [PubMed: 30783251]
15. Zhang J. et al. Germline mutations in predisposition genes in pediatric cancer. N. Engl. J. Med. 373, 2336–2346 (2015). [PubMed: 26580448]
16. Huang K-L et al. Pathogenic germline variants in 10,389 adult cancers. Cell 173, 355–370.e14 (2018). [PubMed: 29625052]
17. Lu C. et al. Patterns and functional implications of rare germline variants across 12 cancer types. Nat. Commun. 6, 10086 (2015). [PubMed: 26689913]
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). [PubMed: 28873162]
19. Jonsson P. et al. Tumour lineage shapes BRCA-mediated phenotypes. Nature 571, 576–579 (2019). [PubMed: 31292550]
20. Cheng DT 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. Genomics 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). [PubMed: 25741868]
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). [PubMed: 9751605]
23. Win AK, Hopper JL & Jenkins MA Association between monoallelic MUTYH mutation and colorectal cancer risk: a meta-regression analysis. Fam Cancer 10, 1–9 (2011).
24. Knudson AG Two genetic hits (more or less) to cancer. Nat. Rev. Cancer 1, 157–162 (2001). [PubMed: 11905807]
25. Nickerson ML et al. Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the Birt-Hogg-Dubé syndrome. Cancer Cell 2, 157–164 (2002). [PubMed: 12204536]
26. Alexandrov LB et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013). [PubMed: 23945592]
27. Polak P. et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat. Genet. 49, 1476–1486 (2017). [PubMed: 28825726]
28. Pilati C. et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J. Pathol. 242, 10–15 (2017). [PubMed: 28127763]

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). [PubMed: 28551381]
30. Scarpa A. et al. Whole-genome landscape of pancreatic neuroendocrine tumours. Nature 543, 65–71 (2017). [PubMed: 28199314]
31. Chakravarty D. et al. Oncokb: A precision oncology knowledge base. JCO Precis. Oncol. 2017, (2017).
32. Le DT et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 357, 409–413 (2017). [PubMed: 28596308]
33. Lynch HT, Shaw MW, Magnuson CW, Larsen AL & Krush AJ Hereditary factors in cancer. Study of two large midwestern kindreds. Arch. Intern. Med. 117, 206–212 (1966). [PubMed: 5901552]
34. Lynch HT & Krush AJ Cancer family "G" revisited: 1895–1970. Cancer 27, 1505–1511 (1971). [PubMed: 5088221]
35. Latham A. et al. Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer. J. Clin. Oncol. 37, 286–295 (2019). [PubMed: 30376427]
36. Møller 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). [PubMed: 28754778]
37. Kinzler KW & Vogelstein B. Landscaping the cancer terrain. Science 280, 1036–1037 (1998). [PubMed: 9616081]
38. Kinzler KW & Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature 386, 761, 763 (1997). [PubMed: 9126728]
39. Baysal BE et al. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. Science 287, 848–851 (2000). [PubMed: 10657297]
40. Karaayvaz-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).
41. Cheng DT 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 HH, Scott SN, Brannon AR, Shah RH & Berger MF Detecting somatic genetic alterations in tumor specimens by exon capture and massively parallel sequencing. J. Vis. Exp. e50710 (2013). doi:10.3791/50710
43. Xin J. et al. High-performance web services for querying gene and variant annotation. Genome Biol. 17, 91 (2016). [PubMed: 27154141]
44. Karczewski KJ et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv (2019). doi:10.1101/531210
45. Landrum MJ et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 44, D862–8 (2016). [PubMed: 26582918]
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). [PubMed: 15256438]
47. Hayward BE et al. Extensive gene conversion at the PMS2 DNA mismatch repair locus. Hum. Mutat. 28, 424–430 (2007). [PubMed: 17253626]
48. Kilpivaara O, Alhopuro P, Vahteristo P, Aaltonen LA & Nevanlinna H. CHEK2 I157T associates with familial and sporadic colorectal cancer. J. Med. Genet. 43, e34 (2006). [PubMed: 16816021]
49. Kilpivaara O. et al. CHEK2 variant I157T may be associated with increased breast cancer risk. Int. J. Cancer 111, 543–547 (2004). [PubMed: 15239132]
50. Apostolou P. & Papasotiriou I. Current perspectives on CHEK2 mutations in breast cancer. Breast Cancer (Dove Med Press) 9, 331–335 (2017). [PubMed: 28553140]
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). [PubMed: 31444830]

52. Alam NA 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). [PubMed: 12761039]

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). [PubMed: 23576677]

55. Benjamini Y. & Hochberg Y. Controlling the false discovery rate: A Practical and powerful approach to multiple testing. J.Roy.Statist.Soc. (1995).

56. Bielski CM et al. Widespread selection for oncogenic mutant allele imbalance in cancer. Cancer Cell 34, 852–862.e4 (2018). [PubMed: 30393068]

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). [PubMed: 28481359]

58. Chang MT et al. Accelerating discovery of functional mutant alleles in cancer. Cancer Discov. 8, 174–183 (2018). [PubMed: 29247016]

59. Vogelstein B. et al. Cancer genome landscapes. Science 339, 1546–1558 (2013). [PubMed: 23539594]

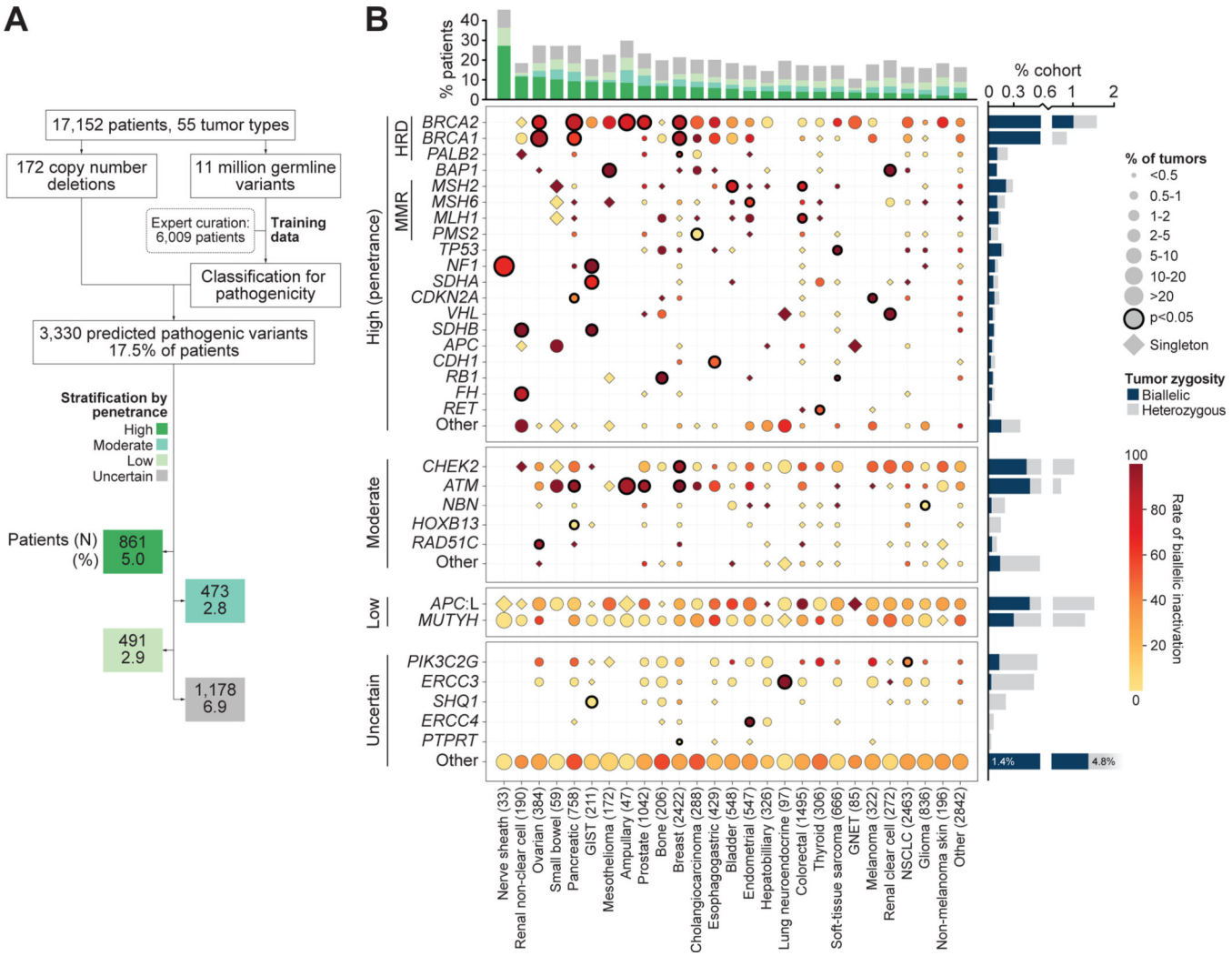


Fig. 1: Germline pathogenicity in prospectively characterized advanced cancers.

A) Study schema describing the identification and pathogenicity assessment of germline carriers and the number and percent of affected patients. 3,158 pathogenic germline variants and 172 germline copy number deletions were identified. 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. Size of the circle denotes prevalence of mutations in each listed gene within a given cancer type. Color of the circles corresponds to the rate of biallelic inactivation of the gene within the cancer type. Circles with dark outlines indicate statistically significant gene-tumor type associations after correcting for multiple hypothesis testing (q<0.15); diamonds denote singletons. Bar chart at right denotes overall prevalence of each gene broken down by biallelic and heterozygous rates. Bar chart at the top denotes prevalence of mutations associated with high, moderate, low and uncertain penetrance within individual tumor types. *APC*:L reflects the I1307K variant associated with low penetrance.

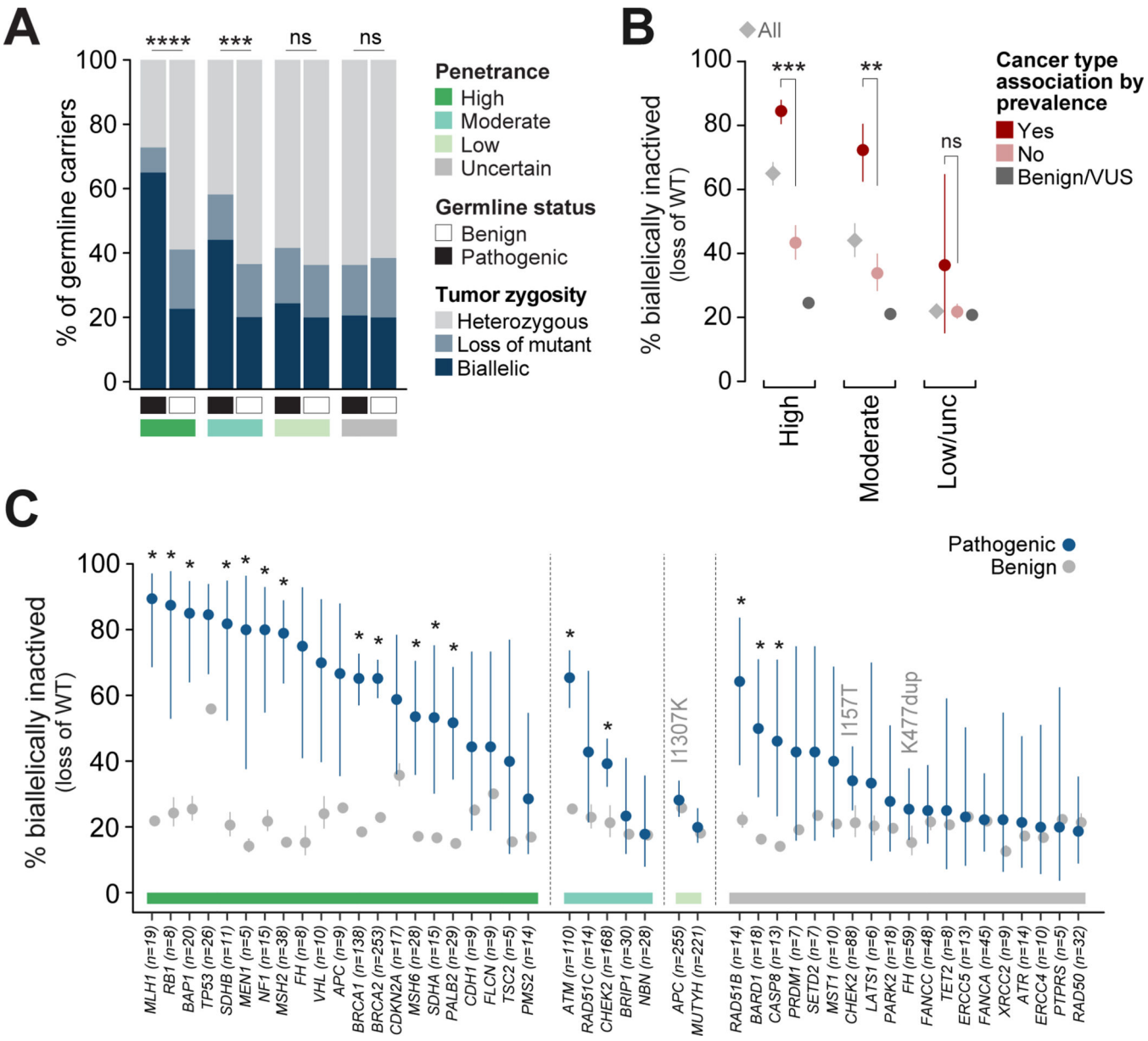


Fig. 2: Penetrance and tumor lineage drive selection for somatic biallelic inactivation.

A) The rate of somatic biallelic inactivation targeting either pathogenic or known/presumed benign variants by penetrance level (legend indicates germline status, penetrance class, and somatic zygosity). **** $P = 2.8\times10^{-125}$; *** $P = 5.7\times10^{-27}$; n.s. not significant; logistic regression accounting for cancer type, specimen type (primary/metastasis), genomic instability and sex. **B)** By penetrance level, the rate of somatic biallelic inactivation of the germline allele 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 (714 pathogenic; 45,302 benign), moderate (354; 16,185) and low/uncertain (1,353; 128,686) penetrance variants. Error bars are 95% binomial confidence intervals (CIs). *** $P = 2.6\times10^{-21}$; ** $P = 5\times10^{-5}$; n.s. not significant; logistic regression accounting for cancer type, specimen type (primary/metastasis), genomic

instability and sex. **C)** The rate of somatic biallelic inactivation of germline pathogenic variants by gene and penetrance (for genes with five or more germline pathogenic variants and a biallelic inactivation rate >15%). Shown are the fraction with somatic biallelic inactivation among carriers of pathogenic variants (blue) and benign variants (gray) within the same gene. Error bars are 95% binomial CIs. * Benjamini-Hochberg adjusted $P < 0.05$; logistic regression accounting for cancer type, specimen type (primary/metastasis) and genomic instability.

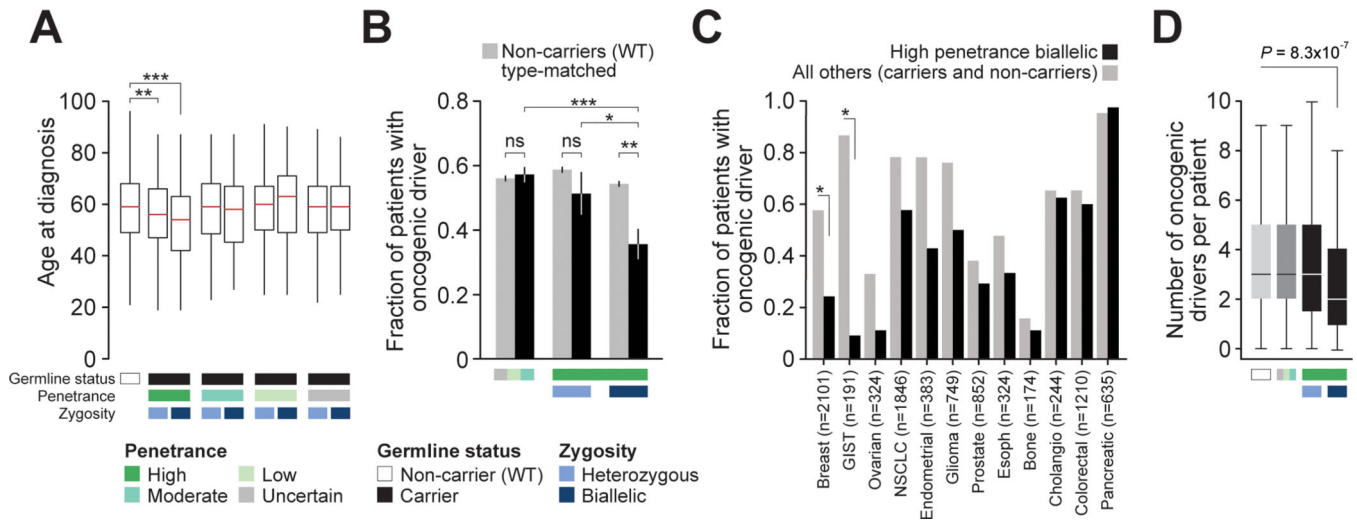


Fig. 3: Phenotypic and evolutionary consequences of biallelic inactivation high penetrance alleles.

A) Age of cancer diagnosis was younger only in carriers of pathogenic variants associated with high penetrance, and younger still in those with biallelic inactivation. Pediatric patients (younger than 18 years of age) were excluded from this analysis. Data shown for 662 high, 389 moderate, 412 low, 927 uncertain penetrance carriers and 11,950 germline WT patients. *** $P = 6\times10^{-14}$; ** $P = 1.9\times10^{-5}$; linear regression adjusting for cancer type, specimen type (primary/metastasis), genomic instability (see Methods) 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 diagnosis. The upper and lower whiskers extend up to 1.5 * IQR (interquartile range) above and below the upper and lower hinges, respectively. **B)** The fraction of patients with tumors harboring gain-of-function driver alterations (oncogenic mutations, focal CNAs, or fusions) as a function of their germline status (see legend, U/L/M are uncertain, low, or moderate penetrance alleles). To control for tumor type, tumors from non-carrier (germline WT) patients were randomly sampled with matched distributions of cancer types (see Methods). * $P = 2.5\times10^{-4}$; ** $P = 2\times10^{-7}$; *** $P = 1.4\times10^{-14}$; ns, not significant; two-sided chi-squared test among indicated classes. Data shown for 11,428 non-carriers, 611 high and 1609 moderate/low/uncertain penetrance carriers. Error bars are 95% binomial CIs. **C)** The fraction of patients with somatic oncogenic alterations by cancer type in carriers of germline alleles that are biallelic or not in the corresponding cancers (black and gray, respectively). * Bonferroni adjusted $P < 0.05$; Breast, $P = 1.3\times10^{-10}$; GIST, $P = 1.1\times10^{-6}$; two-sided Fisher's exact test. In parentheses are total number of patients for each cancer type. Tumor types with fewer than five biallelic carriers were excluded from this analysis. **D)** The number of somatic drivers including both gain-of-function and loss-of-function alterations per patient by germline status (same as in panel B); two-way analysis of variance (ANOVA). Data shown is same as panel B. For the boxplots, the center line is the median, the lower and the upper hinges represent the first and third quartiles for the number of somatic drivers. The upper and lower whiskers extend up to 1.5 * IQR (interquartile range) above and below the upper and lower hinges, respectively.

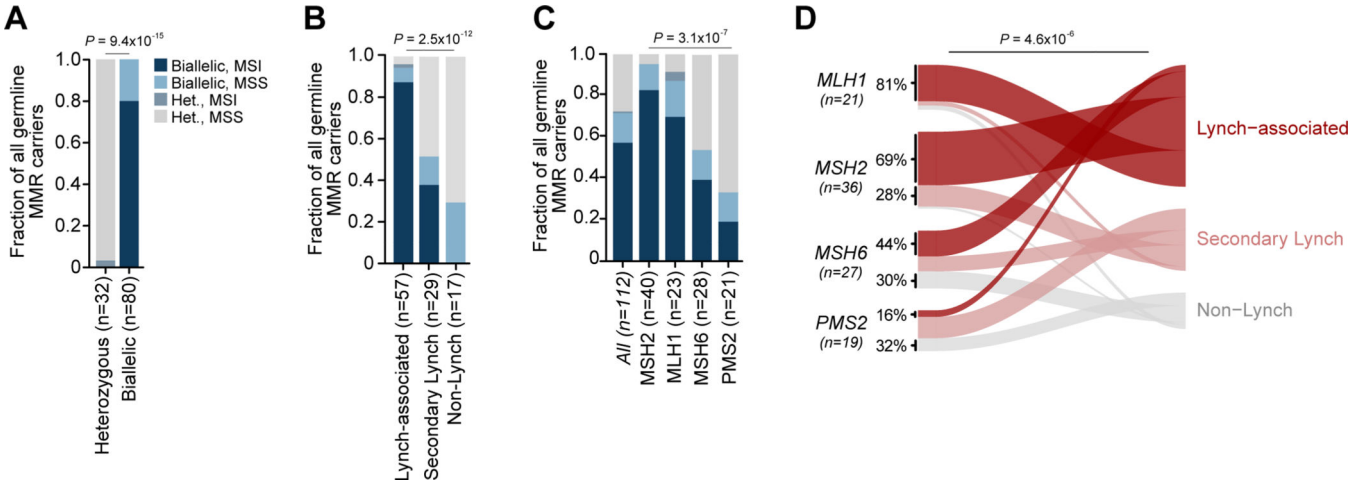


Fig. 4: Lineage and zygosity-dependent somatic phenotypes of germline mismatch repair (MMR) carriers. A)

The proportion of germline MMR carriers (pathogenic *MSH2*, *MSH6*, *MLH1*, or *PMS2*) that have tumors of microsatellite instability (MSI) compared to those that are microsatellite stable (MSS) as a function of zygosity status; two-sided Fisher's exact test. **B**) The presence of the MSI phenotype in germline MMR carriers (defined as in panel A) that are biallelic in their corresponding tumor grouped by cancer type, either Lynch-associated (colon, rectum, endometrium, ovary, upper tract urothelial, stomach, small intestine and adrenocortical) or Secondary Lynch associated (hepatobiliary, pancreas, prostate brain tumors, renal cell and sarcomas) or Not-Lynch associated (breast, melanoma, lung, and thyroid). two-sided chi-square test for trend in proportions. **C)** The proportion of germline MMR carriers (defined as in panel A) by gene altered grouped by zygosity (chi-square test for trend in proportions) and MSI phenotype ($P = 0.13$, two-sided chi-square test for trend in proportions for proportion MSI out of biallelic). **D)** The proportion of germline MMR carriers by gene altered within each cancer type group, either Lynch-associated, Secondary Lynch or Not-Lynch associated. two-sided chi-square test for trend in proportions.

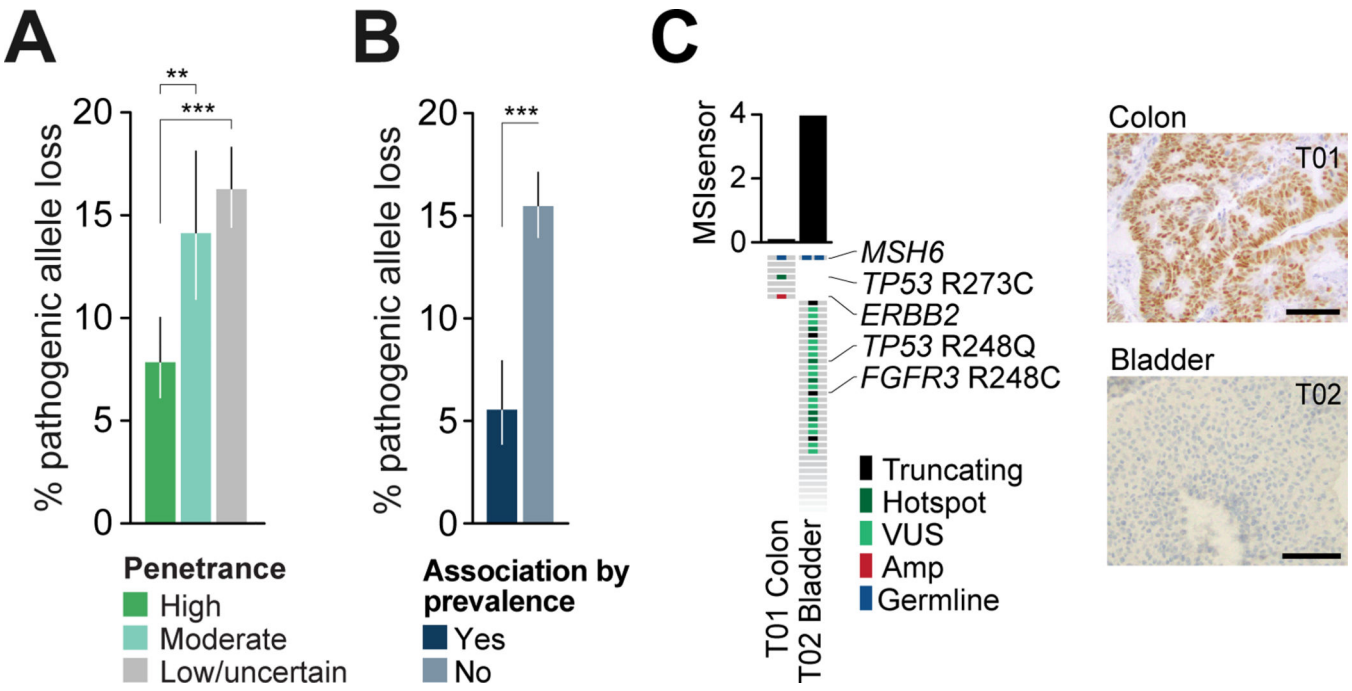


Fig. 5: Dispensability of germline pathogenicity in tumorigenesis.

A) The percent of cases harboring germline alleles associated with either high (n=714), moderate (n=354), or low/uncertain (n=1,353) penetrance that are deleted in the incumbent clone of the resulting tumor by LOH targeting the mutant allele is shown. *** $P = 4.2\times10^{-8}$, ** $P = 0.0016$; two-sided chi-squared test for trend in proportions. Error bars represent the 95% binomial CIs. **B)** The rate of somatic loss of the germline pathogenic allele based on whether the affected tumor was of cancer types associated with the germline allele by prevalence (n=480) or not (n=1,941) is shown. Error bars are 95% binomial CIs; *** $P = 1.5\times10^{-9}$, two-sided Fisher's exact test. **C)** Two independent cancers, one colorectal and one bladder primary (T01 and T02 respectively) diagnosed in an *MSH6* K218* germline carrier in which only the bladder cancer had biallelic inactivation of the germline allele and harbored the corresponding MSI phenotype (genomics, left). At right is immunohistochemistry confirming MSH6 was null in only the bladder cancer (T02). Scale bars represent 100 μm.

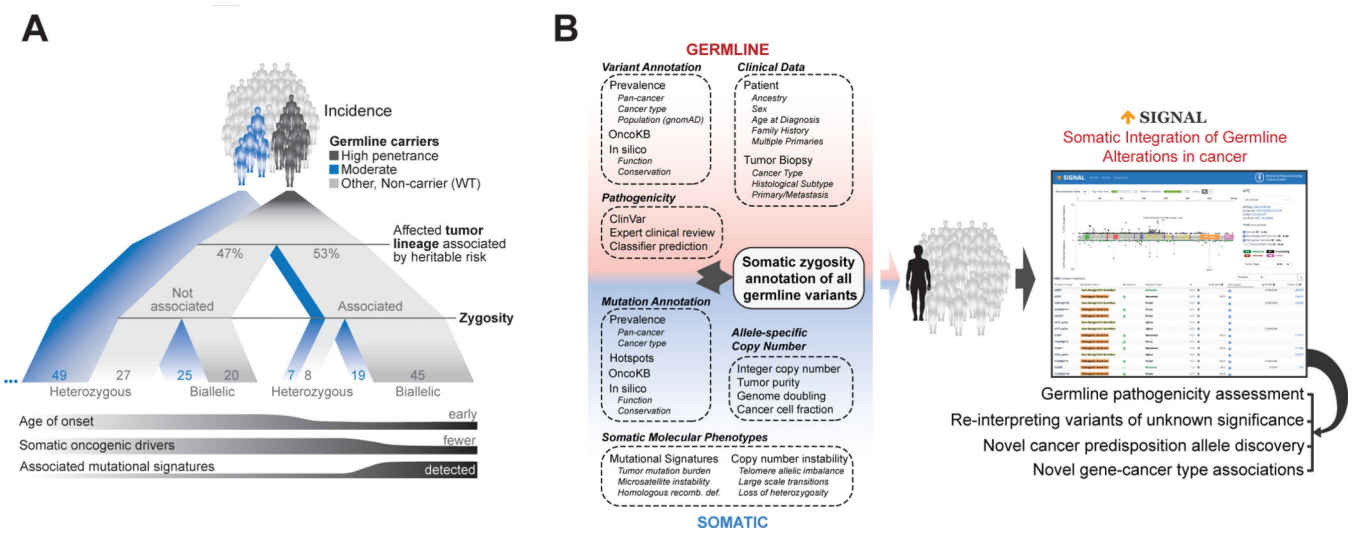


Fig. 6: Integration of germline and somatic tumor profiling.

A) Model of dependence of tumorigenesis on germline pathogenicity. In carriers of pathogenic variants associated with high penetrance (5.1% overall, dark gray), tumors of associated cancer types that acquire somatic biallelic inactivation were most common and exhibited somatic and clinical hallmarks of dependence on the germline allele such as early age of onset and fewer number of required somatic oncogenic drivers. In contrast, tumors with pathogenic germline variants in the majority of moderate penetrance genes or in non-associated cancer types appeared to be unrelated cancers that lacked these dependencies. **B)** Integrated analysis of somatic and germline alterations for clinical assessment. Interpretation of pathogenic germline alterations is informed by tumor-specific zygosity and associated somatic and clinical features to provide a holistic view of a patient's cancer. SignalDB captures the rates of germline variants, zygosity, and co-occurring somatic alterations population-wide to aid the clinical interpretation of future variants and facilitate broader discovery.