



Published in final edited form as:

*Hum Mutat.* 2018 November ; 39(11): 1542–1552. doi:10.1002/humu.23640.

## Integrating Somatic Variant Data and Biomarkers for Germline Variant Classification in Cancer Predisposition Genes

Michael F. Walsh<sup>1,\*</sup>, Deborah I. Ritter<sup>2,\*</sup>, Chimene Kesserwan<sup>3,\*</sup>, Dmitriy Sonkin<sup>4</sup>, Debyani Chakravarty<sup>1</sup>, Elizabeth Chao<sup>5</sup>, Rajarshi Ghosh<sup>2</sup>, Yelena Kemel<sup>1</sup>, Gang Wu<sup>3</sup>, Kristy Lee<sup>6</sup>, Shashi Kulkarni<sup>2</sup>, Dale Hedges<sup>3</sup>, Diana Mandelker<sup>1</sup>, Ozge Ceyhan-Birsoy<sup>1</sup>, Minjie Luo<sup>7</sup>, Michael Drazer<sup>8</sup>, Liying Zhang<sup>1</sup>, Kenneth Offit<sup>1</sup>, and Sharon E. Plon<sup>2</sup>

<sup>1</sup>Memorial Sloan Kettering Cancer Center

<sup>2</sup>Baylor College of Medicine

<sup>3</sup>St. Jude Children's Hospital

<sup>4</sup>National Cancer Institute

<sup>5</sup>University of California, Irvine

<sup>6</sup>University of North Carolina

<sup>7</sup>Children's Hospital of Philadelphia

<sup>8</sup>University of Chicago.

### Abstract

In its landmark paper about Standards and Guidelines for the Interpretation of Sequence Variants, the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) did not address how to use tumor data when assessing the pathogenicity of germline variants. The Clinical Genome Resource (ClinGen) established a multidisciplinary working group, the Germline/Somatic Variant Subcommittee (GSVS) with this focus. The GSVS implemented a survey to determine current practices of integrating somatic data when classifying germline variants in cancer predisposition genes. The GSVS then reviewed and analyzed available resources of relevant somatic data, and performed integrative germline variant curation exercises. The committee determined that somatic hotspots could be systematically integrated into moderate evidence of pathogenicity (PM1). Tumor RNA sequencing data showing altered splicing may be considered as strong evidence in support of germline pathogenicity (PVS1) and tumor phenotypic features such as mutational signatures be considered supporting evidence of pathogenicity (PP4). However, at present, somatic data such as focal loss of heterozygosity and mutations occurring on the alternative allele are not recommended to be systematically integrated, instead, incorporation of this type of data should take place under the advisement of multidisciplinary cancer center tumor-normal sequencing boards.

\*Drs. Walsh, Ritter and Kesserwan should be considered joint first author.

## Keywords

Somatic; germline; PM1; variant interpretation; hotspot; PP4; signature

---

## INTRODUCTION

Predominant questions driving both oncologists and cancer geneticists are under what circumstances information generated from genomic testing can be used to inform therapeutics or impact plans for risk reduction and prevention strategies (Kauff et al., 2002; Lima et al., 2018; Villani et al., 2016). Currently, major cancer centers perform parallel sequencing of tumor and matched normal (germline or constitutional) DNA, which can simultaneously reveal information about therapeutic options and cancer predisposition (Dela Cruz et al., 2016; Mody et al., 2015; Oberg et al., 2016; Parsons et al., 2016; Walsh et al., 2014). Two large-scale studies, one pediatric and one adult, have used similar methodology integrating somatic data to augment germline variant calling (Huang et al., 2018; Zhang et al., 2015).

In contrast to research studies, distinct classification guidelines have been published for clinical reporting of tumor variants, including a recent publication from the Association for Molecular Pathology (AMP), germline variants from the American College of Medical Genetics and AMP, and guidance for germline variant reporting in cancer genes (Chakravarty et al., 2017; Li et al., 2017; Richards et al., 2015; Plon et al. 2008). However, these classification schemes do not offer guidance about using variant data from germline sequencing when interpreting tumor variation, or conversely using cancer somatic data in the context of reporting germline variants. Specifically, the evidence codes described in ACMG/AMP germline classification do not reference the use of somatic data, yet a natural set of questions arise as to the appropriate use of existing somatic data to aid germline interpretation (Figure 1). We describe here a standardized approach to incorporating somatic data for germline interpretation (Figure 2).

## METHODS

The Clinical Genome Resource (ClinGen, [clinicalgenome.org](http://clinicalgenome.org)) Hereditary Cancer Clinical Domain Working Group convened the Germline/Somatic Variant Subcommittee (GSVS) to understand current practices and make recommendations for the use of somatic data as a criterion for germline variant interpretation for hereditary cancer genes. The GSVS is comprised of oncologists, geneticists, molecular pathologists, molecular geneticists, bioinformatics specialists, computational biologists, and laboratory directors. We first designed a survey to assess current practices for the use of somatic data for germline variant classification (Supp\_Materials). The design of this survey was based on expert opinion within GSVS to target four main evidence types; 1) mutational hotspots, 2) tumor RNA sequencing (RNA-seq) data, 3) loss of heterozygosity (LOH), and 4) tumor phenotypic characteristics such as signature, mutational burden and microsatellite instability. The survey was sent to 40 molecular laboratories that sequence both tumor and germline samples, and

included in the AMP listserv and [genetests.org](http://genetests.org). The Memorial Sloan Kettering Cancer Center Institutional Review Board (IRB) approved this study.

Next, we identified peer-reviewed and publicly available somatic data sources that could be used for germline variant classification based upon these four evidence types and then defined best practices and limitations for the use of specific datasets. Following this, we selected ACMG/AMP evidence codes that could best incorporate somatic data, using existing codes whenever possible. Finally, we selected 45 peer reviewed rare variants with associated tumor data from The Cancer Genome Atlas (TCGA) experience that were classified as VUS in ClinVar and classified these variants using the approaches we developed. Rules were discussed as a group, applied by individual biocurators and then analyzed as a group for consistency and agreement in the evidence code usage.

## RESULTS

### Laboratory and variant interpretation survey

Of the 21 respondents to the survey, 16 (76.2%) reported following the ACMG/AMP standards and guidelines for classification of germline variants (Richards et al., 2015). Integrating somatic and germline data was not performed by most laboratories (18/21 (85.7%)). When positing that should both tumor and normal sequencing data be available for a given patient 13/18 (72%) respondents reported that they would use somatic data for classifying germline variants. Attribution of somatic data regarding the strength of the evidence was variable; 9/18 (50%) reported considering somatic data at a supporting level, 2/18 (11.1%) at moderate level, and 7/18 (38.9%) at strong level. In contrast, 9/18 (50%) respondents reported they would not incorporate LOH or tumor copy number alteration in companion tumor normal sequencing as evidence for pathogenicity of a germline variant. Additional questions were asked regarding the types of data laboratories use (or would use) as complimentary diagnostics or biomarkers aiding in interpreting germline variants in cancer predisposition genes. Data elements included: immunohistochemistry, telomere lengths, and chromosomal breakage studies 13/18 (72%), RNA sequencing data 15/18 (83.3%) for splicing variants in cancer predisposition genes of uncertain significance, and therapeutic responses 12/18 (66.7%) (Supp\_Materials). As anticipated, the survey identified no consensus on the applications, sources, weights of evidence, or use of somatic data including LOH or mutational hotspots. However, we used these survey results to direct discussion on the working group calls, which aided not only in understanding current usage but prioritizing somatic variant data elements for germline interpretation.

### Publicly available somatic data resources

The survey showed that there was heterogeneity in the sources of somatic data and usage of those sources. Somatic data sources can include any of the following: variant data from paired tumor / normal sequencing, institutional resources such as case registries, local or gene/disease focused resources, public or institutional data, and proprietary datasets. To address this heterogeneity, we reviewed data sources and compiled a list of peer-reviewed resources identified from the survey and review of the literature with a focus on those, which are publicly available (Supp\_Materials). The cBioPortal for Cancer Genomics, originally

developed at Memorial Sloan Kettering Cancer Center (MSK) and currently developed and maintained by a multi-institutional team including The Dana Farber Cancer Institute, Princess Maxima Cancer Center in Toronto, Children's Hospital of Philadelphia, The Hyve in the Netherlands and Bilkent University in Ankara, Turkey (<http://www.cbioportal.org>, Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012), is a publicly available resource that stores somatic variants from 224 cancer studies including those from multiple TCGA projects and the National Moonshot Cancer initiative project GENIE (Genomics Evidence Neoplasia Information Exchange) (Consortium, 2017) (Chang et al., 2016; Chang et al., 2018)), and all hotspots identified by this algorithm are also publicly available at the website [cancerhotspots.org](http://cancerhotspots.org).

### Integration of somatic mutational hotspots (PM1 evidence code)

Here, we provide an overview of cancer mutational hotspots, as this was the somatic data element the committee considered for incorporation into germline variant interpretation for cancer predisposition given the availability of a curated database of somatic hotspots (Chang et al., 2016; Chang et al., 2018). A somatic mutational hotspot was defined as a single amino acid position in a protein-coding gene that is mutated more frequently than would be expected in the absence of selection (Chang, et al., 2016). While the exact definition varies depending on the approach used to calculate hotspots (Chang et al., 2016; Chang et al., 2018; Huang et al., 2018), the methodology assigns a statistical significance to the recurrence of mutation at a given amino acid corrected for the background mutational rate of the position, gene, and sample both within and across cancer types in the affected cohort. Somatic mutational hotspots are therefore not common germline benign variants in a population. For this analysis, we focused on ~1100 mutational hotspots from a recent analysis of the sequencing data from ~25,000 diverse primary and metastatic human cancers available both in the [cancerhotspots.org](http://cancerhotspots.org) portal as well as in the cBioPortal (Chakravarty et al., 2017; Chang et al., 2018; Seiler et al., 2018; Yang et al., 2018). Each codon may be mutated to one or more alternative amino acids primarily in a single cancer type or across many cancer types (Figure 3). Chang et al. defines hotspots at sites with a Q-value <0.1 as statistically significant with a false discovery rate <10%. We identified this data type for potential integration into the ACMG/AMP guidelines Moderate (PM1) evidence category [(PM1 = "located in a mutational hotspot and/or critical and well-established functional domain (e.g. active site of an enzyme) without benign variation" (Richards et al., 2015)]. Notably, hotspots defined per Chang et al, 2016 & 2018 are consistent with the odds ratios described recently of moderate evidence at 4.3:1 and strong evidence at 18.7:1 odds by Tavtigian et al (2018). We piloted this approach on the tumor suppressors *TP53*, *VHL*, *DNMT3A*, *BRCA2*, *PTEN*, *ATM* and the oncogene *PTPN11*.

### Using TP53 somatic hotspots to understand PM1 specifications

The tumor suppressor gene *TP53* is the most commonly mutated gene in cancer and one of the most well studied genes in hereditary and sporadic cancer (Baker, Kinzler, & Vogelstein, 2003; F. P. Li & Fraumeni, 1969). The hotspot database [cancerhotspots.org](http://cancerhotspots.org) includes 120 *TP53* codons for a total of 622 *TP53* variants (Chang et al., 2016; Chang et al., 2018). The TP53 transactivation assay is an established functional assay for *TP53* variants (Kato et al., 2003; Monti et al., 2007; Petitjean et al., 2007). We reviewed TP53 transactivation activity

from the International Agency for Research on Cancer (IARC) *TP53* database (Bouaoun et al., 2016) for all *TP53* amino acid (AA) positions identified as hotspots in [cancerhotspots.org](http://cancerhotspots.org). We identified two key areas in which hotspot data should receive additional considerations for proper use: (1) alternative AA and frequency at the hotspot residue and (2) conflation of variant effects -- missense versus truncating/null -- at hotspot residues. While most missense and null variants at *TP53* mutational hotspots residues are functionally inactive as determined by the IARC reported transactivation assays, a minority of missense variants within mutational hotspot residues may retain functional activity and produce different degrees of transactivation activity (Supp\_Table S3). For example, the *TP53* Ile255 position is considered a mutational hotspot residue ([cancerhotspots.org](http://cancerhotspots.org)) when substituted with a Phe, Asn, Ser or Thr; all 4 mutations have been found in cancer and all lack functional p53 transactivation activity. However, a Val substitution at position 255 is not considered a cancer hotspot and consistently, the p.Ile255Val substitution does not show a significant decrease in transactivation activity relative to wild type p53 (Kato et al., 2003). As the hotspot calculation in [cancerhotspots.org](http://cancerhotspots.org) is summarized over all contributing amino acids, this underscores the importance of considering AA identity and frequency at each hotspot residue when incorporating hotspot data into variant pathogenicity interpretation.

Additionally, somatic hotspot usage should avoid conflation of missense and predicted truncating/null data by careful consideration of the gene functional context (tumor suppressor genes vs oncogene) and variant type (predicted truncating/null vs. activating in tumor suppressor versus oncogenes respectively). Currently in [cancerhotspots.org](http://cancerhotspots.org), predicted truncating/null and missense variants are combined for hotspot Q-values. Hotspots in tumor suppressor genes may be driven almost entirely by predicted truncating/null mutations while missense mutations at that position may not have functional impact. This is exemplified by NP\_001119584.1(*TP53*):p.(Glu294\*) a predicted truncation/null hotspot, whereas missense mutations at this same AA residue 294 maintains intact p53 transactivation ability comparable to wild type p53. Another example is with NP\_001120982.1(*APC*):p.Gln1378\* where all variants in [cancerhotspots.org](http://cancerhotspots.org) identified as hotspots are truncating/null. We therefore suggest any use of hotspot data should separate data derived from missense changes from predicted null or truncating variants, especially if they are predicted to lead to nonsense mediated decay.

### Analysis of VHL at Somatic Hotspots to understand PM1 specifications

In the recent publication from Tavgigian et al, the authors derive the moderate level of evidence as a 4:3:1 odds of pathogenicity (2018). To test if somatic hotspot data could uphold this odds ratio, we searched for *VHL* somatic variants in [cancerhotspots.org](http://cancerhotspots.org) that have received germline classification in ClinVar (Supp\_Table). Out of 69 hotspot variants, we identified a total of 34 *VHL* germline variants in ClinVar, of which 30 variants were considered pathogenic or likely pathogenic and 4 as VUS, Conflicting interpretation, Likely Benign, or Benign. This generated an odds of pathogenicity of ~7.5:1, which is higher than moderate evidence odds ratios (4.3:1) from Tavgigian et al but less than strong evidence odds ratios (18.7:1). To be conservative, we chose to maintain consistency with moderate evidence presented by Richards et al and Tavgigian et al (2015, 2018), noting that the clinical significance field in ClinVar can vary in the number and consensus of laboratories

contributing, and we did not further analyze data based on the number of contributing laboratories.

### Specifying hotspot data for somatic data integration (PM1 evidence code)

For the most applicable and reproducible somatic data inclusion, we propose using the **PM1 evidence code** if the variant AA at the hotspot codon in question has a sample count in [cancerhotspots.org](http://cancerhotspots.org) that is equal or greater than 10 at that codon. If the variant AA has a sample count that is greater than 1, and less than 10 (from 2–9), a PM1\_supporting evidence code is recommended (Figure 4). In addition, when evaluating the somatic hotspot data, it is strongly suggested to consider the following points:

1) *Hereditary Cancer Context*: Somatic hotspot data is recommended for use only in the context of a germline variant being interpreted with regard to cancer predisposition, as some cancer susceptibility genes are also associated with non-cancer syndromes. 2) *Germline and Somatic Hotspots*: If an amino acid residue is a mutational hotspot both at the germline and somatic level, then only the germline information should be used to fulfill the evidence code. Even in cases where somatic evidence leads to a stronger conclusion for the PM1 code (ex. PM1 for somatic and PM1\_Supporting for germline), germline evidence is prioritized for a germline interpretation. 3) *Variant Type and Database Composition and Cancer Spectrum*: To identify statistically significant or recurrent hotspots in cancer, the [cancerhotspots.org](http://cancerhotspots.org) algorithm used a population-scale cohort of tumor samples of various cancer types. It must be noted however, that the majority of patient tumor samples used were solid tumors in adults and while 41 tumor types were used in their analysis some pediatric tumor types were not included (Chang et al. 2016). Consequently, the cancers contributing to the hotspot residue may not fully represent the cancer(s) associated with the patient's hereditary cancer predisposition syndrome. We suggest careful review of the cancer types and variants at the hotspot residue and to consider this information within the context of the patient's cancer phenotype and/or family history of cancer.

### Variant classification examples using PM1 evidence code specification

To identify variants for this pilot, we used the intersection of 2671 variants from 226 genes in [cancerhotspots.org](http://cancerhotspots.org) with those present in the ClinVar variant\_summary.txt file (downloaded on 4/12/2018, Supp. Tables S3 and S4). We queried germline variant interpretations submitted to ClinVar using the Genome Reference Consortium Human Build 37 GRCh37 (hg19). We focused on 109 variants with Uncertain Significance and 97 variants with Conflicting Interpretations of Pathogenicity. We used the ACMG five-tier system classification: Pathogenic, Likely Pathogenic, Uncertain Significance (also described as Variant of Uncertain Significance or VUS, below), Likely Benign and Benign. Here we provide examples of germline variant analysis applying PM1 evidence (Supp\_Materials).

NM\_000546.5(*TP53*):c.374C>T;p.(Thr125Met): This variant had Conflicting Interpretations of Pathogenicity (VUS and Likely Pathogenic) in ClinVar (ID 183748). This variant was evaluated in the ClinGen Variant Curation Interface (VCI) as VUS with the following evidence codes: PS3\_Supporting, PM2, PP3, and PP4. In [cancerhotspots.org](http://cancerhotspots.org) 16/30 samples



were c.374C>T; p. (Thr125Met). The PM1 evidence code was applied, leading to the upgrade of this variant to Likely Pathogenic.

NM\_000546.5(*TP53*):c.845G>T;p.(Arg282Leu): This variant had conflicting interpretations of pathogenicity -- VUS and Likely Pathogenic -- in ClinVar (ID 182938). In the ClinGen VCI, we evaluated this variant as Likely Pathogenic with PM2, PM5, and PP3 evidence codes. In [cancerhotspots.org](http://cancerhotspots.org), the overwhelming number of samples harbored the Trp amino acid change (201 / 219 samples) whereas the L amino acid change was seen in just 1 sample (bowel cancer). However, the R282 codon is a known germline hotspot, and applying the PM1 evidence code further supports the interpretation of Likely Pathogenic and helps clarify prior discordant VUS classifications (Baugh et al., 2018). See Discussion section for variants that are both a germline and somatic hotspot.

NM\_000546.5(*TP53*): c.542G>A; p. (Arg181His): This variant had conflicting interpretations of pathogenicity (Likely Pathogenic and Pathogenic) in ClinVar (ID 142320) but was reported by a commercial lab as a VUS (7/8/17). Using the ClinGen VCI, we interpreted this variant as a VUS based on PP3, PP4, PM2\_Supporting, and PS3\_Moderate codes. In [cancerhotspots.org](http://cancerhotspots.org), 9/26 samples had the Arg181His variant. When the PM1\_Supporting evidence code for somatic hotspot data is applied, the variant becomes Likely Pathogenic.

NM\_00314.6(*PTEN*): c.395G>A; p.(Gly132Asp): This variant had conflicting interpretations of pathogenicity with two VUS interpretations and one Likely Pathogenic in ClinVar(ID:92822). Using the ClinGen VCI, we interpreted this variant as VUS based on the following: PP2, PP3, PP4, and PM2 codes. In [cancerhotspots.org](http://cancerhotspots.org), mutations of Gly882Asp at this site were found in 8/17 samples. Adding PM1\_Supporting somatic hotspot makes the interpretation Likely Pathogenic.

NM\_000551.3(*VHL*): c.452T>C; p.(Ile151Thr): This variant had conflicting interpretations of pathogenicity with one submission as VUS and one as Likely Pathogenic in ClinVar(ID: 428803). Using the ClinGen VCI, we interpreted this variant as VUS based on the following: PP3, PP4, and PM2 codes. In [cancerhotspots.org](http://cancerhotspots.org), Ile882Thr mutations were found in 2/6 samples. Adding PM1\_Supporting somatic hotspot does not change the interpretation, leaving it as VUS.

NM\_002834.4(*PTPN11*): c.215C>T; p.(Ala72Val): This variant had conflicting interpretations of pathogenicity with one submission as VUS, four submissions as Likely Pathogenic and one submission as Pathogenic in ClinVar (ID:41443). Using the ClinGen VCI, we interpreted this variant as Likely Pathogenic based on the following: PP2, PP3, PP4, PM2 and PS3. In [cancerhotspots.org](http://cancerhotspots.org), mutations of Ala882Val were found in 6/18 samples. Adding PM1\_Supporting somatic hotspot code makes the interpretation Pathogenic.

NM\_022552.4(*DNMT3A*): c.2645G>A; p. (Arg882His): This variant had conflicting interpretations of pathogenicity -- VUS and Pathogenic in ClinVar (ID:375881). Using the ClinGen VCI, we interpreted this variant as Likely Pathogenic based on the following: PM2\_Supporting (given that the variant is not completely absent in the Genome Aggregation

Database but is present at a very low allele count), PP3, PP2 and PS3. In [cancerhotspots.org](https://cancerhotspots.org), mutations of R882H were found in 28/39 samples. With the addition of PM1 using somatic hotspot data, this variant remains Likely Pathogenic.

NM\_000051.3(*ATM*): c.1009C>T (p. Arg337Cys): This variant is interpreted as a VUS in ClinVar (ID: 127327). This was identified in a Caucasian, 75-year-old, woman with invasive ductal carcinoma of the breast, metastatic to bone and lymph nodes, presenting with a paternal aunt with postmenopausal breast cancer, and a paternal cousin with breast cancer at ~30 years of age. Using the ClinGen Variant Curation Interface (VCI), the variant was classified as a VUS using PM2\_Supporting and PP3. In [cancerhotspots.org](https://cancerhotspots.org), the p.Arg337Cys variant was found in 31/40 samples. However, with the addition of PM1 the variant remained a VUS.

#### Additional Relevant Somatic Data Types for Interpreting Germline Variants

In addition to specifying systematic use of somatic hotspot data in the PM1 evidence code for the interpretation of heritable cancer variants, we considered potential evidence codes for usage of other relevant somatic data types such as tumor signatures, chromothripsis, mutational burden, microsatellite instability (MSI), tumor RNA-seq data and LOH. Here, we summarize the potential uses of somatic data in the PP4 and PVS1 evidence codes, and LOH data.

#### Integration of RNA-Seq tumor data (PVS1 evidence code)

The current definition of the **PVS1 evidence code** from Richards et al is “null variant (nonsense, frameshift, canonical  $\pm 1$  or 2 splice sites, initiation codon, single or multi exon deletion) in a gene where LOF is a known mechanism of disease”. Analysis of RNAseq data from tumors may provide insights into germline variants and provide further evidence for the application of codes such as PVS1. For example, the use of tumor-derived RNA sequencing data can determine whether a canonical or non-canonical predicted splice site variant results in abnormal cDNA isoforms that are associated with disruptions in splicing (Figure 5). Variants can occur at splice sites or at an intron / exon junction and splice disruption can lead to a truncated protein or nonsense-mediated decay (Seiler et al., 2018; Yang et al., 2018; Zhang et al., 2015). In Zhang et al, tumor RNAseq was used to support likely functional germline splicing disruption caused by a predicted splice variant in *ATM*, as the RNAseq displayed marked loss of read counts in exons 3’ of the splice site (see Figure S4 in Zhang et al., 2015). Further insight for usage of PVS1 evidence code has been detailed recently (see paper by Tayoun et al. in this same issue). Following their PVS1 decision tree, the splice site in the example from Zhang et al. would receive the full PVS1 evidence code, whereas without tumor RNA-seq data of the germline variant to confirm nonsense mediated decay, the maximum it could receive would be PVS1\_Strong (downgrade from Very Strong to Strong).

#### Loss of Heterozygosity is not to be used routinely for germline variant interpretation

We initially proposed systematic incorporation of LOH data to ACMG/AMP germline classification given that a finding of tumor LOH could increase the functional impact of



germline variants in cancer predisposition syndromes (Kanchi et al., 2014). However, there are two main barriers to standardizing the use of LOH evidence for routine germline classification: (1) Definition of the length of LOH that would constitute a specific, or focal, loss and (2) Variation between and a range of qualities of somatic LOH calling algorithms from next generation sequencing data. Both barriers could lead to misinterpretations of the pathogenicity of the remaining allele after LOH events. For example, results from tumor mutation panels often do not provide enough data to distinguish whether the LOH event is centered on the gene containing the germline variant in question or instead represents LOH across an entire chromosome segment or arm containing hundreds of genes. Also problematic is the variation in somatic LOH calling software, and performance, particularly in the exome sequencing setting. We considered many examples relevant for LOH and highlight two such examples here in *ATM* and *BRCA2* genes seen in 2 patients with breast cancer and brain cancer, respectively. These 2 variants, designated NM\_000051.3(*ATM*): c.8071C>T p.(Arg2691Cys) and NM\_000059.3 (*BRCA2*): c.6058G>A p.(Glu2020Lys) are classified as uncertain significance in ClinVar, and reported to have somatic focal LOH data (Lu et al., 2015). For the *ATM* p.(Arg2691Cys) variant, supporting LOH data was only available in one reported case, and for *BRCA2* p.(Glu2020Lys), the tumor type was atypical. Based on this and the technical issues with LOH, we recommend that LOH data not be used routinely for integration into germline variant classification.

However, using LOH on a case-by-case basis may be accepted when there is sufficient expertise in both germline and somatic variant and LOH calling, with documentation of LOH deemed focal by experts, consistency of tumor type between the germline variant and cancer type, and any reported LOH data is seen in >1 case. In such cases, it may be possible to incorporate LOH data (Figure 5) to substantiate PVS1 evidence code. In these cases, we suggest that LOH data is used to increase the PVS1 evidence code found by following the decision tree in Tayoun et al (in this same issue). For example, following the PVS1 decision tree, a nonsense variant in a tumor suppressor gene, lacking evidence of nonsense mediated decay, nonsense variants in that exon are rare in the population, and less than 10% of the protein is truncated would receive a PVS1\_Moderate score. With the addition of somatic LOH (as deemed focal by experts), the germline variant could be assigned PVS1\_Strong. For LOH not deemed focal by experts, we do not recommend the use of PVS1 evidence code.

### **Integration of Biomarkers for Somatic Support of Germline Variant Interpretation (PP4 evidence code)**

In addition to specifying use of somatic hotspot data in the PM1 evidence code for the interpretation of heritable cancer variants, we considered potential evidence codes for usage of other relevant somatic data types such as tumor signatures, chromothripsis, mutational burden, and microsatellite instability (MSI). The **PP4 evidence code** (“Patient’s phenotype or family history is highly specific for a disease with a single genetic etiology”) can potentially incorporate these four types of evidence (Figure 5 & Summary Box 1). Large-scale sequencing studies have provided mutational signatures sensitive to detecting germline Hereditary Breast and Ovarian Cancer (HBOC) and constitutional mismatch repair (MMR) deficient patients (Alexandrov et al., 2013; Campbell et al., 2017). For example, tumor

signatures 3 and 6 (as defined by Alexandrov et al., 2013) when detected through tumor and matched normal sequencing could be considered as supporting evidence for phenotypes associated with the following genes: *BRCA2* (signature 3) and *MLH1/MSH2/MSH6/PMS2* (signature 6). The signatures of mutational processes in cancer are publicly available online ([cancer.sanger.ac.uk](http://cancer.sanger.ac.uk)). While these signatures may indicate functional consequence of an alteration in one of these genes, they are not specific to individual germline variants, and therefore we suggest supporting phenotypic evidence (PP4). In addition to tumor signature 13, characterized by large-scale tumor rearrangements and chromothripsis may be used as PP4 evidence, as they are indicative of a germline predisposition such as medulloblastoma and germline *TP53* mutations (Grobner et al., 2018; Rausch et al., 2012). Thus, in instances where tumor signatures are available, in addition to molecular sequencing, this should be used as supporting evidence.

The total number of variants detected in a tumor may signify an underlying predisposition (Bouffet et al., 2016; Campbell et al., 2017; Le et al., 2015; Shlien et al., 2015). However, there is little consensus on the definition of hypermutation, which leads to differences in application and interpretation (Campbell et al., 2017; Pritchard et al., 2016). Moreover, the cause of hypermutation -- such as hyper-methylation and silencing of *MLH1* or specific therapeutic interventions -- should be considered in addition to constitutional MMR deficiency or Lynch syndrome (Campbell et al., 2017; Dudley, Lin, Le, & Eshleman, 2016; van Thuijl et al., 2015). When hypermutation of tumors is not a result of treatment, it can support a phenotypic picture consistent with a germline predisposition in mismatch repair. Microsatellite instability is an additional assay providing a means to identify patients with predisposition to cancer. Currently, however, this is still a non-specific marker with similar caveats to mutational burden (Campbell et al., 2017). These data should ideally be considered in a patient specific context and under the advisement of a tumor/germline review committee.

## DISCUSSION

The use of somatic data has been a valuable resource for the classification of germline variants in cancer predisposition genes in large research studies (Huang et al., 2018; Zhang et al., 2015). As indicated by our survey of laboratories, most laboratories would use somatic data to interpret germline variants, if somatic data accompanied germline data for a given patient. The incorporation of somatic data to germline variant evaluation requires expertise, quality somatic data and an understanding of the complexities of tumor data when considering its use for clinical germline classification. We initially discussed incorporating multiple types of somatic data by adding additional evidence codes to the ACMG criteria for the classification of germline variants, such as a new supporting evidence tag if the variant is present in a somatic database. However, a validation exercise of 27 established germline variants with integration of LOH as a moderate tag revealed that 8/27 variants would change from an established Benign or Likely Benign classification to Uncertain and 5/27 would move from Uncertain to Likely Pathogenic (Supp\_Table). Ultimately, we determined that creating new ACMG evidence codes at this time could not be done with scientific rigor given the available data. While we considered somatic data broadly, we propose an overall conservative application of somatic data in line with the ACMG/AMP Standards and

Guidelines. Mutational hotspots in cancer genes have been defined and are reproducible based on gene size, tumor type and aggregated sequencing studies (Chang et al., 2016; Chang et al., 2018). Thus, we, with the guidance and approval of the ClinGen Sequence Variant Interpretation (SVI) WG which oversees specifications to ACMG criteria, recommend the use of statistically significant somatic hotspot data using the PM1 evidence code for germline classification of variants in hereditary cancer genes. However, it is important to note that hotspot definitions will undergo iterative changes as more tumors are sequenced. An example is *TP53*, where hotspots at AA residues 157 and 158 were identified only after the tumor sample set expanded (11,119 versus 24,592 tumor samples) (Chang et al., 2018). This emphasizes the abundance of data necessary to discover hotspots and highlights the potential that some true hotspots may be missed. Additionally, careful consideration should be given to cases when a residue is both a germline and a somatic hotspot, such as the *TP53* variant curation example of arginine codon 282. Although currently we recommend using the germline hotspot information instead, it is worthwhile to consider using both sources of information in an additive approach, with potential increase in the strength of the evaluation based on hotspot evidence from both germline and somatic sources. As more somatic hotspots are identified and as germline variant curation efforts further define germline hotspots as well, it is likely that we will identify more sites with both somatic and germline hotspot evidence. In a future analysis, we may seek to additively integrate this information and [www.clinicalgenome.org](http://www.clinicalgenome.org) will host any updates to these recommendations.

We also sought initially to systematically integrate LOH data, but reconsidered due to the many caveats which now limit its standardized use in clinical classification of germline variants. Notably, multiple research-based tumor boards are using LOH data to aid in classifying rare variants, as seen in the Pediatric Cancer Genome Project (PCGP) Germline Study, and three The Cancer Genome Atlas (TCGA) efforts (Huang et al., 2018; Kanchi et al., 2014; Lu et al., 2015; Seiler et al., 2018; Walsh et al., 2014; Yang et al., 2018; Zhang et al., 2015). Each of these studies assessed variants in the context of patient specific paired sequencing. However, loss of the wild type allele in the tumor should be combined with other evidence for pathogenicity such as functional biomarkers or characteristic genomic signatures (Davies et al., 2017; Alexandrov et al., 2013; Riaz et al., 2017). As nuances for LOH are further specified, such as critical size for focal LOH, and LOH computational calling programs on whole exome improve, we expect this data element to be more meaningfully used (Abkevich et al., 2012; Koboldt et al., 2012). In addition, our committee aims to periodically re-evaluate LOH evidence and create a new recommendation when warranted.

Evidence generated as part of comprehensive cancer sequencing and evaluation can help guide and strengthen support for interpretations of pathogenicity in germline cancer predisposition genes. These data elements are beneficial when analyzed on a case-by-case basis, considering the many different parameters of these types of experiments and analytic platforms. As tumor profiles have been analyzed in parallel to germline sequencing for patients with cancer predisposition, characteristics of the tumor profile including tumor signatures, mutational burdens and microsatellite instability can support well-known phenotypic classifications. We suggest PP4 evidence code in these cases. Furthermore,

RNA-seq from the tumor, when available, can provide the use of the very strong evidence (PVS1) code to validate germline splicing variants, although caution must be taken if the tumor has experienced significant structural variation as they may ablate the mutated allele or make RNA-seq interpretation challenging.

Evaluating somatic data in the context of hereditary cancer has proved to be a challenging task. When the Germline / Somatic Variant Subcommittee formed, we considered several questions (Figure 1), in part drawn from our own experience evaluating hereditary cancer patients and incorporating concurrent tumor sequencing. Given the sizable number of clinical laboratories considering integrating somatic data (62%), and the diverse sources of potential somatic data elaborated by the survey, it is apparent, at present, it will be a major challenge to integrate additional relevant somatic data due to inconsistency in use, interpretation, and training and in some cases methodology (LOH). The challenge can be overcome by improved understanding and increased use of tumor sequencing data, and enhanced training in tumor boards and centers that prioritize germline and tumor somatic sequencing for cancer care. We have summarized the data types, limitations and uses in a table (Table 1). We acknowledge that these recommendations may evolve, as more data emerges and as we consider additional data types, such as somatic epigenetic silencing in a recent paper from Park et al (2018). The ClinGen Germline / Somatic Variant Subcommittee aims to continue to assess somatic data for standardized incorporation into germline variant evaluation, and conversely, to explore whether germline data can provide insight into somatic variant evaluations (Li et al., 2017). It may be the case that once tumor and germline sequencing is standard of care, a merged variant interpretation guideline that draws information from both sources to guide interpretation of tumor and germline variants will provide the greatest clinical utility.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

The ClinGen GVSC would like to acknowledge Barry Taylor, PhD for his expertise and explanations in defining strengths and limitations of mutational hotspot data. We would also like to acknowledge the experts and bio curators on our working group for their dedicated work in variant curation (<https://www.clinicalgenome.org/working-groups/clinical-domain/hereditary-cancer-clinical-domain-working-group/somatic-germline-variant-curation-group/>). Michael F. Walsh receives research support from the Crawford Genomics Fund, the Corning Fund, the Niehaus Center for Inherited Cancer Genomics and the V Foundation for Cancer Research. Sharon E. Plon serves on the Baylor College of Medicine Scientific Advisory Board.

GRANT NUMBERS: The ClinGen consortium is also funded by the National Human Genome Research Institute and the Eunice Kennedy Shriver National Institute of Child Health and Human Development through contracts 1U41HG006834-01A1, 1U01HG007437-01, 1U01HG007436-01, 1U01HG006487-01, and U41HG009650, U41HG009649

## References

Abida W, Armenia J, Gopalan A, Brennan R, Walsh M, Barron D, ... Scher HI (2017). Prospective Genomic Profiling of Prostate Cancer Across Disease States Reveals Germline and Somatic Alterations That May Affect Clinical Decision Making. *JCO Precis Oncol*, 2017. doi:10.1200/PO.17.00029

- Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, ... Lanchbury JS (2012). Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer*, 107(10), 1776–82. doi: 10.1038/bjc.2012.451. [PubMed: 23047548]
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, ... Stratton MR (2013). Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415–421. doi: 10.1038/nature12477 [PubMed: 23945592]
- Baker SJ, Kinzler KW, & Vogelstein B (2003). Knudson's hypothesis and the TP53 revolution. *Genes Chromosomes Cancer*, 38(4), 329. doi:10.1002/gcc.10249 [PubMed: 14566852]
- Ballinger ML, Best A, Mai PL, Khincha PP, Loud JT, Peters JA, ... Savage SA (2017). Baseline Surveillance in Li-Fraumeni Syndrome Using Whole-Body Magnetic Resonance Imaging: A Meta-analysis. *JAMA Oncol*, 3(12), 1634–1639. doi:10.1001/jamaoncol.2017.1968 [PubMed: 28772291]
- Baugh EH, Ke H, Levine AJ, Bonneau RA, & Chan CS (2018). Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death and Differentiation*, 25(1), 154–160. doi: 10.1038/cdd.2017.180. [PubMed: 29099487]
- Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, & Olivier M (2016). TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum Mutat*, 37(9), 865–876. doi:10.1002/humu.23035 [PubMed: 27328919]
- Bouffet E, Larouche V, Campbell BB, Merico D, de Borja R, Aronson M, ... Tabori U (2016). Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency. *J Clin Oncol*, 34(19), 2206–2211. doi:10.1200/JCO.2016.66.6552 [PubMed: 27001570]
- Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, ... Shlien A (2017). Comprehensive Analysis of Hypermutation in Human Cancer. *Cell*, 171(5), 1042–1056 e1010. doi: 10.1016/j.cell.2017.09.048 [PubMed: 29056344]
- Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, ... Schultz N (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol*, 2017. doi:10.1200/PO.17.00011
- Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, ... Taylor BS (2016). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol*, 34(2), 155–163. doi:10.1038/nbt.3391 [PubMed: 26619011]
- Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P, ... Taylor BS (2018). Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discov*, 8(2), 174–183. doi:10.1158/2159-8290.CD-17-0321 [PubMed: 29247016]
- Claes K, Dahan K, Tejpar S, De Paepe A, Bonduelle M, Abramowicz M, ... Kartheuser A (2011). The genetics of familial adenomatous polyposis (FAP) and MutYH-associated polyposis (MAP). *Acta Gastroenterol Belg*, 74(3), 421–426. [PubMed: 22103048]
- Consortium APG (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov*, 7(8), 818–831. doi:10.1158/2159-8290.CD-17-0151 [PubMed: 28572459]
- Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, ... Nik-Zainal S (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med*, 23(4), 517–525. doi:10.1038/nm.4292 [PubMed: 28288110]
- Dela Cruz FS, Diolaiti D, Turk AT, Rainey AR, Ambesi-Impiombato A, Andrews SJ, ... Kung AL (2016). A case study of an integrative genomic and experimental therapeutic approach for rare tumors: identification of vulnerabilities in a pediatric poorly differentiated carcinoma. *Genome Med*, 8(1), 116. doi:10.1186/s13073-016-0366-0 [PubMed: 27799065]
- Dudley JC, Lin MT, Le DT, & Eshleman JR (2016). Microsatellite Instability as a Biomarker for PD-1 Blockade. *Clin Cancer Res*, 22(4), 813–820. doi:10.1158/1078-0432.CCR-15-1678 [PubMed: 26880610]
- Furutani E, & Shimamura A (2017). Germline Genetic Predisposition to Hematologic Malignancy. *J Clin Oncol*, 35(9), 1018–1028. doi:10.1200/JCO.2016.70.8644 [PubMed: 28297620]
- Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, ... Pfister SM (2018). The landscape of genomic alterations across childhood cancers. *Nature*, 555(7696), 321–327. doi:10.1038/nature25480 [PubMed: 29489754]

- Harris MH, DuBois SG, Glade Bender JL, Kim A, Crompton BD, Parker E, ... Janeway KA (2016). Multicenter Feasibility Study of Tumor Molecular Profiling to Inform Therapeutic Decisions in Advanced Pediatric Solid Tumors: The Individualized Cancer Therapy (iCat) Study. *JAMA Oncol*. doi:10.1001/jamaoncol.2015.5689
- Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, ... Ding L (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*, 173(2), 355–370 e314. doi:10.1016/j.cell.2018.03.039 [PubMed: 29625052]
- Jasperson KW, Patel SG, & Ahnen DJ (1993). APC-Associated Polyposis Conditions. In Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, & Amemiya A (Eds.), *GeneReviews*(R) Seattle (WA).
- Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, ... Ding L (2014). Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun*, 5, 3156. doi: 10.1038/ncomms4156 [PubMed: 24448499]
- Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, & Ishioka C (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A*, 100(14), 8424–8429. doi: 10.1073/pnas.1431692100 [PubMed: 12826609]
- Kauff ND, Satagopan JM, Robson ME, Scheuer L, Hensley M, Hudis CA, ... Offit K (2002). Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation. *N Engl J Med*, 346(21), 1609–1615. doi:10.1056/NEJMoa020119 [PubMed: 12023992]
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, ..., Wilson RK (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*, 22(3):568–76. doi: 10.1101/gr.129684.111 [PubMed: 22300766]
- Kratz CP, Achatz MI, Brugieres L, Frebourg T, Garber JE, Greer MC, ... Malkin D (2017). Cancer Screening Recommendations for Individuals with Li-Fraumeni Syndrome. *Clin Cancer Res*, 23(11), e38–e45. doi:10.1158/1078-0432.CCR-17-0408 [PubMed: 28572266]
- Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, ... Diaz LA, Jr. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*, 372(26), 2509–2520. doi: 10.1056/NEJMoa1500596 [PubMed: 26028255]
- Li FP, & Fraumeni JF, Jr. (1969). Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med*, 71(4), 747–752. [PubMed: 5360287]
- Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, ... Nikiforova MN (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*, 19(1), 4–23. doi: 10.1016/j.jmoldx.2016.10.002 [PubMed: 27993330]
- Lima AD, Alves VR, Rocha AR, Martinhago AC, Martinhago C, Donadio N, ... Gebrim LH (2018). Preimplantation genetic diagnosis for a patient with multiple endocrine neoplasia type 1: case report. *JBRA Assist Reprod*, 22(1), 67–70. doi:10.5935/1518-0557.20180010 [PubMed: 29266898]
- Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, ... Ding L (2015). Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun*, 6, 10086. doi:10.1038/ncomms10086 [PubMed: 26689913]
- Mandelker D, Zhang L, Kemel Y, Stadler ZK, Joseph V, Zehir A, ... Offit K (2017). 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(9), 825–835. doi: 10.1001/jama.2017.11137 [PubMed: 28873162]
- Mody RJ, Wu YM, Lonigro RJ, Cao X, Roychowdhury S, Vats P, ... Chinnaiyan AM (2015). Integrative Clinical Sequencing in the Management of Refractory or Relapsed Cancer in Youth. *JAMA*, 314(9), 913–925. doi:10.1001/jama.2015.10080 [PubMed: 26325560]
- Monti P, Ciribilli Y, Jordan J, Menichini P, Umbach DM, Resnick MA, ... Fronza G (2007). Transcriptional functionality of germ line p53 mutants influences cancer phenotype. *Clin Cancer Res*, 13(13), 3789–3795. doi:10.1158/1078-0432.CCR-06-2545 [PubMed: 17606709]



- Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, ... Kung AL (2016). Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. *Genome Med*, 8(1), 133. doi:10.1186/s13073-016-0389-6 [PubMed: 28007021]
- Park S, Supek F, Lehner B (2018). Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. *Nat Commun*, 9(1), 2601. doi: 10.1038/s41467-018-04900-7 [PubMed: 29973584]
- Parsons DW, Roy A, Yang YP, Wang T, Scollon S, Bergstrom K, ... Plon SE (2016). Clinical genomics for children with solid tumors: Current realities and future opportunities. *Clinical Cancer Research*, 22. doi:Doi 10.1158/1557-3265.Pmsclingen15-Ia16
- Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, & Olivier M (2007). Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat*, 28(6), 622–629. doi:10.1002/humu.20495 [PubMed: 17311302]
- Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, ... IARC Unclassified Genetic Variants Working Group. (2008). Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human Mutation*, 29(11), 1282–1291. doi: 10.1002/humu.20880 [PubMed: 18951446]
- Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, ... Nelson PS (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med*, 375(5), 443–453. doi:10.1056/NEJMoa1603144 [PubMed: 27433846]
- Rausch T, Jones DT, Zapatka M, Stutz AM, Zichner T, Weischenfeldt J, ... Korbel JO (2012). Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell*, 148(1–2), 59–71. doi:10.1016/j.cell.2011.12.013 [PubMed: 22265402]
- Riaz N, Bleuca P, Lim RS, Shen R, Higginson DS, Weinhold N, ... Reis-Filho JS (2017). Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. *Nat Commun*, 8(1), 857. doi:10.1038/s41467-017-00921-w [PubMed: 29021619]
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, ... Committee, A. L. Q. A. (2015). 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(5), 405–424. doi:10.1038/gim.2015.30 [PubMed: 25741868]
- Schrader KA, Cheng DT, Joseph V, Prasad M, Walsh M, Zehir A, ... Robson M (2016). Germline Variants in Targeted Tumor Sequencing Using Matched Normal DNA. *JAMA Oncol*, 2(1), 104–111. doi:10.1001/jamaoncol.2015.5208 [PubMed: 26556299]
- Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, ... Yu L (2018). Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep*, 23(1), 282–296 e284. doi:10.1016/j.celrep.2018.01.088 [PubMed: 29617667]
- Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, ... Biallelic Mismatch Repair Deficiency, C. (2015). Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermuted cancers. *Nat Genet*, 47(3), 257–262. doi: 10.1038/ng.3202 [PubMed: 25642631]
- Sucheston-Campbell LE, Clay A, McCarthy PL, Zhu Q, Preus L, Pasquini M, ... Hahn T (2015). Identification and utilization of donor and recipient genetic variants to predict survival after HCT: are we ready for primetime? *Curr Hematol Malig Rep*, 10(1), 45–58. doi:10.1007/s11899-014-0246-x [PubMed: 25700678]
- Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, & Biesecker LG (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. 10.1038/gim.2017.210
- Tawana K, & Fitzgibbon J (1993). CEBPA-Associated Familial Acute Myeloid Leukemia (AML). In Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, & Amemiya A (Eds.), *GeneReviews*(R) Seattle (WA).
- van Thuijl HF, Mazor T, Johnson BE, Fouse SD, Aihara K, Hong C, ... Costello JF (2015). Evolution of DNA repair defects during malignant progression of low-grade gliomas after temozolomide

treatment. *Acta Neuropathol*, 129(4), 597–607. doi:10.1007/s00401-015-1403-6 [PubMed: 25724300]

Verlinsky Y, Rechitsky S, Verlinsky O, Chistokhina A, Sharapova T, Masciangelo C, ... Kuliev A (2002). Preimplantation diagnosis for neurofibromatosis. *Reprod Biomed Online*, 4(3), 218–222. [PubMed: 12709270]

Villani A, Shore A, Wasserman JD, Stephens D, Kim RH, Druker H, ... Malkin D (2016). Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. *Lancet Oncol*, 17(9), 1295–1305. doi:10.1016/S1470-2045(16)30249-2 [PubMed: 27501770]

Villani A, Tabori U, Schiffman J, Shlien A, Beyene J, Druker H, ... Malkin D (2011). Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study. *Lancet Oncol*, 12(6), 559–567. doi:10.1016/S1470-2045(11)70119-X [PubMed: 21601526]

Walsh M, Wu G, Edmonson M, Gruber TA, Easton J, Yergeau D, ... Downing JR (2014). Incidence of Germline Mutations in Cancer-Predisposition Genes in Children with Hematologic Malignancies: a Report from the Pediatric Cancer Genome Project. *Blood*, 124(21).

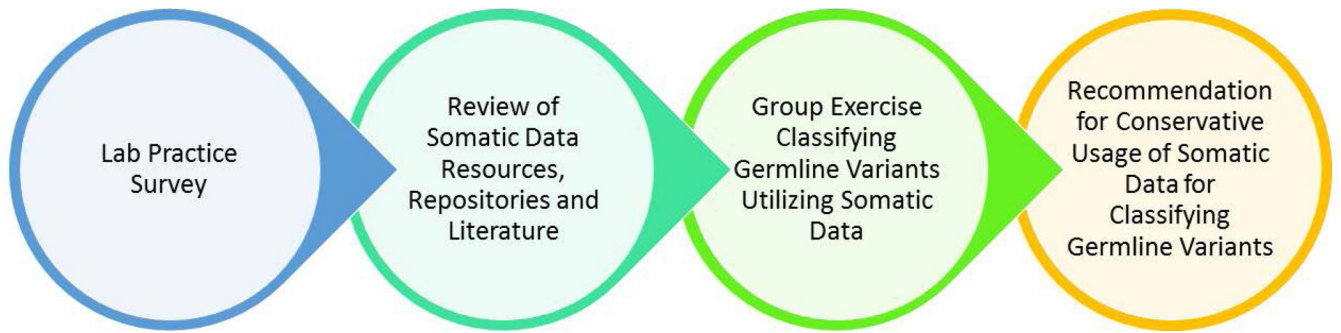
Yang C, Jairam S, Amoroso KA, Robson ME, Walsh MF, & Zhang L (2018). Characterization of a novel germline BRCA1 splice variant, c.5332+4delA. *Breast Cancer Res Treat*, 168(2), 543–550. doi:10.1007/s10549-017-4595-8 [PubMed: 29185120]

Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, ... Downing JR (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*, 373(24), 2336–2346. doi: 10.1056/NEJMoa1508054 [PubMed: 26580448]

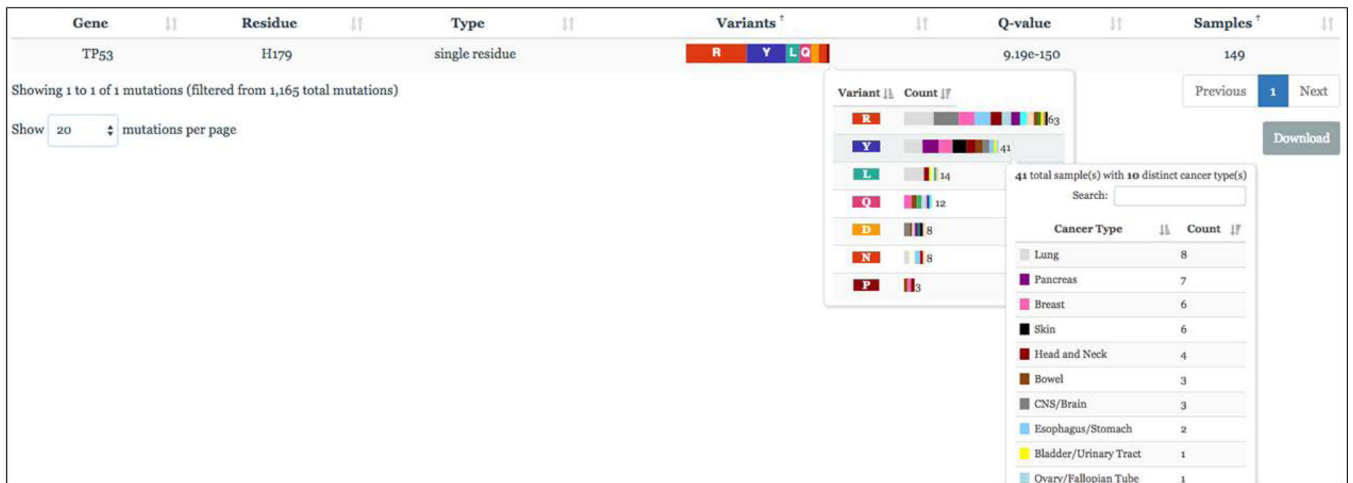
## Check List of Questions When Considering Somatic Data for Aiding in Classifying Germline Cancer Predisposition Gene Variants

- Was there a second hit in the tumor?
- Was there loss of heterozygosity?
- Did the patient have a high mutational burden or Microsatellite Instability high sensor?
- Was staining for gene relevant protein present or absent?
- Was a tumor signature observed?
- Does RNA tumor data support either over or under expression for oncogenes or tumor suppressor gene of interest?
- Does the patient's tumor correspond with the recognized spectrum of tumors for a given predisposition syndrome?

**Figure 1:**  
Checklist of Questions When Considering Somatic Data for Aiding in Classifying Germline Cancer Predisposition Gene Variants

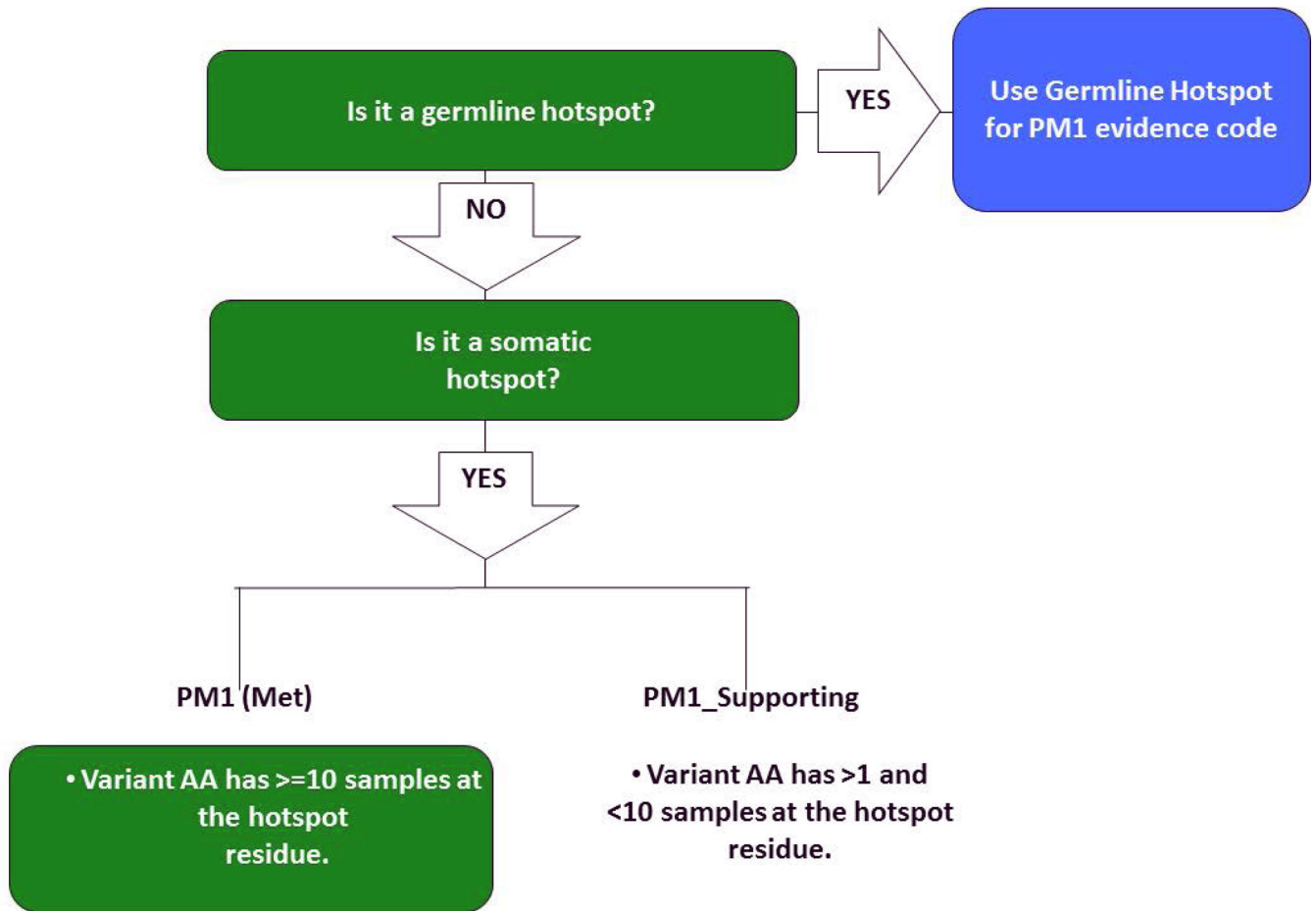


**Figure 2:**  
Process for considering the use of somatic variant data in classifying germline variants



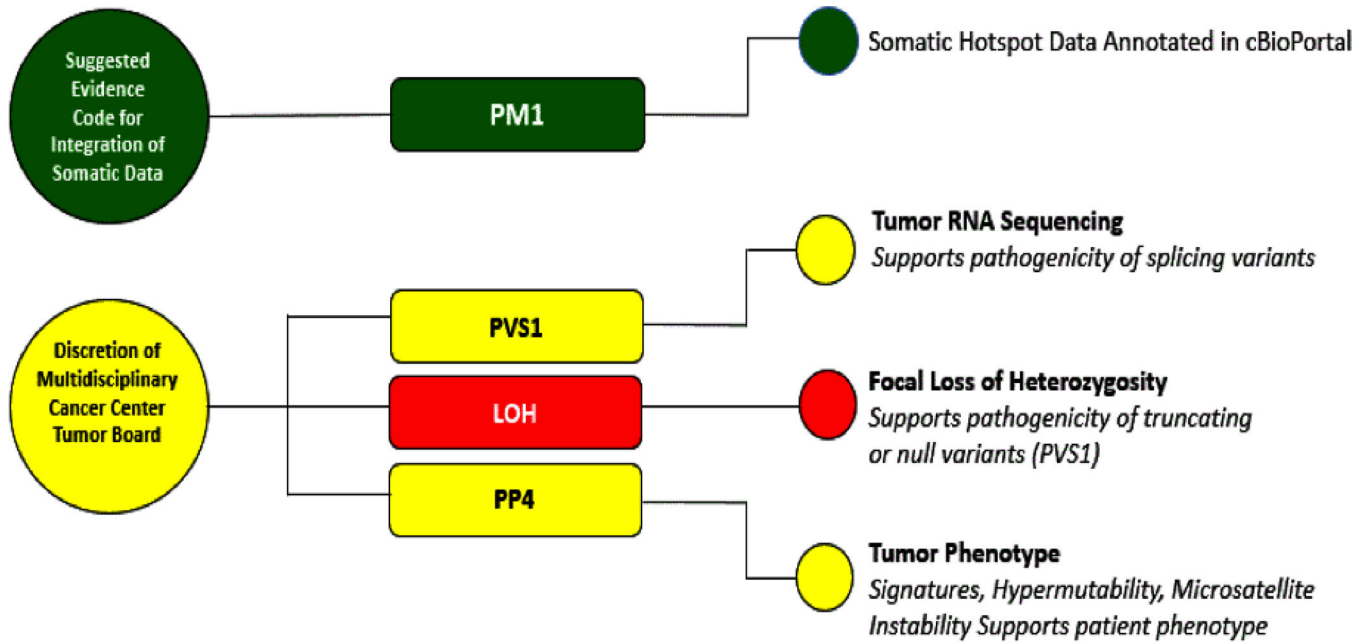
**Figure 3:**

Example of TP53 histidine codon 179 in [cancerhotspots.org](http://cancerhotspots.org). (Far Right) The residue has 149 samples in the dataset, and seven amino acid variants represented by 3, 8, 8, 12, 14, 41, and 63 samples (middle box). Any AA with  $\geq 10$  samples will be PM1 (R, Y, L, Q), while AA with  $>1$  (and  $<10$ ) will be PM1\_Supporting (D, N, P). Each AA additionally displays the number of samples derived from a selection of cancer types (right box).



**Figure 4:**  
PM1 Rule Specifications: Flow chart used to determine if somatic data should be used for PM1, and how it should be applied.





**Figure 5:** Integration of Somatic Data for Germline Variant Classification. Green: Evidence Code for general use and application; Yellow: Evidence codes should only be considered under guidance of multidisciplinary tumor board, as this requires patient specific integrated data and institutional norms; Red: Additional caution needed with LOH to ensure focal loss and relevant germline variant on retained allele.

Table 1:

Types of somatic data and limitations of somatic data usage for annotating germline variants in cancer predisposition genes. LOH: loss of heterozygosity.

Somatic Data Type	Utility	Limitations				Subject to Modification with increasing Metadata
		Defined by Field with Standard Reference	Availability of Assay	Tissue Dependent	Confounding Factors for Interpretation	
Loss of Heterozygosity	Evidence supporting second hit and consistency with Knudson Two--Hit Hypothesis	No	Algorithms to define assay dependent and a definition is not accepted regarding size of LOH	Requires tumor tissue an different Labs have variable requirements pertaining to amount and type	Certain tumor types are difficult to determine, i.e. hypodiploid leukemia	As more tumors sampled greater clarity will be determined regarding focal vs. diffuse LOH definitions
Cancer Hotspots	Mutations defined based metadata, gene size, and tumor types	Yes; <a href="http://cancerhotspots.com">cancerhotspots.com</a> ; <a href="http://cBioPortal.com">cBioPortal.com</a>	Data can be generated by reference laboratories	Requires tumor tissue and different labs have variable requirements pertaining to amount and type	Not all Cancer Types have been factored into calculation of Hotspots. At present 25,000 tumor types included	Definitions have seen changes in definition based on analysis of 11,000 tumors to 25,000
RNA--Seq	Beneficial in determining significance of splicing variants	No	Many commercial and academic labs; but few offer clinically	Requires tumor tissue and different labs have variable requirements pertaining to amount and type; can also be done on non--tumor tissue	Universal standard for interpretation lacking	Yes
Microsatellite Instability Sensor	Provides evidence consistent with mis--match repair deficiency	Yes (but center dependent)	Available by some clinical labs	Yes	Tumor purity can impact interpretation	Yes
Chromothripsis	Provides evidence consistent with germline	Yes; signature 13	Limited requires WGS/WES	No	Tumor purity	No

Somatic Data Type	Utility	Limitations				Subject to Modification with increasing Metadata
		Defined by Field with Standard Reference	Availability of Assay	Tissue Dependent	Confounding Factors for Interpretation	
	TP53 mutation					
Immunohistochemistry (IHC)	Absence of proteins by IHC provides evidence of genetic abnormality, e.g. mismatch repair genes, SDH complex genes	For some genes, e.g., <i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS2</i> , <i>SDHA</i> , <i>SDHB</i>	Commercially available, expensive	Yes	Preparation and amount of tissue available	No
Cancer Signatures	e.g. 3,6,9, 13 when detected can provide evidence of germline abnormality	Yes	Yes, but requires WGS/WES	No	Expensive, Analytic Expertise	Yes
Second Mutation of Alternate Allele	Evidence supporting second hit and consistency with Knudson Two--Hit Hypothesis	No	yes, but to become complete requires extensive	No	Depending on extent of sequencing second mutation may or may not be missed	Yes
Methylation Testing	Epigenetic silencing may indicate evidence supporting second hit and consistency with Knudson Two--Hit Hypothesis	Yes: brain tumors and sarcomas	Limited	Yes	Tumor purity and handling of sample	Yes

**Summary Box 1:**

**Expanded ACMG/AMP Criteria for Applying Tumor Evidence in the Classification of Variants in Cancer Genes**

**Very strong evidence of pathogenicity**

PVS1 Null variant (nonsense, frameshift, canonical +/-1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease

- The presence of loss of heterozygosity in a tumor, judged to be focal by experts, can increase the strength of a downgraded application of PVS1 by one level

Caveat: intended for the context of a germline variant being interpreted with regard to cancer predisposition

**Moderate evidence of pathogenicity**

PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g. active site of an enzyme) without benign variation

- Variant AA at the hotspot codon in question has a sample count in [cancerhotspots.org](http://cancerhotspots.org) that is equal or greater than 10 at that codon Reduce to PM1\_Supporting if variant AA has a sample count that is greater than 1, and less than 10 (from 2–9)

Caveat: intended for the context of a germline variant being interpreted with regard to cancer predisposition

**Supporting evidence of pathogenicity**

PP4 Patient’s phenotype or family history is highly specific for a disease with a single genetic etiology

- Mutation signatures, with statistically significant association to a single gene or small set of genes, all of which were sequenced in the individual(s) with the variant under investigation (e.g. *BRCA2*: signature 3, *MLH1*, *MSH2*, *MSH6*, *PMS2*: signature 6; *TP53*: signature 13)

Caveat: intended for the context of a germline variant being interpreted with regard to cancer predisposition

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript