ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

The Science and Art of Clinical Genetic Variant Classification and Its Impact on Test Accuracy

Hunter H. Giles,[1] Madhuri R. Hegde,[2,3] Elaine Lyon,[4] Christine M. Stanley,[5,6] Iain D. Kerr,[7] Megan E. Garlapow,[1] and Julie M. Eggington[1]

[1]Center for Genomic Interpretation, Sandy, Utah 84092, USA; email: hgiles@genomicinterpretation.org, mgarlapow@genomicinterpretation.org, jeggington@genomicinterpretation.org

[2]PerkinElmer Genomics, Waltham, Massachusetts 02450, USA; email: madhuri.hegde@perkinelmer.com

[3]Department of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

[4]HudsonAlpha Clinical Services Lab, Huntsville, Alabama 35806, USA; email: elyon@hudsonalpha.org

[5]C2i Genomics, Cambridge, Massachusetts 02139, USA

[6]Variantyx, Framingham, Massachusetts 01701, USA; email: christine.stanley@variantyx.com

[7]Ann Arbor, Michigan 48197, USA

Annu. Rev. Genom. Hum. Genet. 2021. 22:285–307

First published as a Review in Advance on April 26, 2021

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-121620-082709

Keywords

variant classification, DNA mutational analyses, precision medicine, rare diseases, oncology, next-generation sequencing

Abstract

Clinical genetic variant classification science is a growing subspecialty of clinical genetics and genomics. The field's continued improvement is essential for the success of precision medicine in both germline (hereditary) and somatic (oncology) contexts. This review focuses on variant classification for DNA next-generation sequencing tests. We first summarize current limitations in variant discovery and definition, and then describe the current five- and four-tier classification systems outlined in dominant standards and guideline publications for germline and somatic tests, respectively. We then discuss measures of variant classification discordance and the field's bias for positive results, as well as considerations for panel size and population screening in the context of estimates of positive predictive value that

incorporate estimated variant classification imperfections. Finally, we share opinions on the current state of variant classification from some of the authors of the most widely used standards and guideline publications and from other domain experts.

1. INTRODUCTION

Clinical genetic variant classification science is hard. Compared with the vast array of other clinical science specialties, genetic variant classification science is in its infancy. The expanding community of variant classification scientists is progressing through inevitable growing pains while learning from both its successes and its unfortunate failures. Although the science of clinical genetic variant classification still has years to go before developing into a mature clinical science, some major breakthroughs and lessons learned from mentors in older clinical science specialties have yielded initial progress. Variant classification scientists have become a significant strength for supporting diagnoses and have already contributed to the implementation of precision medicine, particularly preventative interventions and targeted therapies. As we learn more and remain focused on raising the bar, clinical genetic variant classification scientists expect to continue contributing to the advancement of diagnostic, predictive, prognostic, and theragnostic medicine.

This nonexhaustive review discusses the legacy of genetic variant classification science and how it affects today's methodologies, primarily in the United States. Commentary from domain experts will help to look both backward and forward, with a focus on opportunities to improve the accuracy of the field in the coming years. Note that at the time of writing, major efforts are underway to revise the societal variant classification guidelines for both germline and somatic methodology. Therefore, some of the issues raised here may be in the process of being resolved when this article is published.

2. GENETIC VARIANTS ARE DEFINED USING A REFERENCE SEQUENCE, SOMETIMES INCORRECTLY

A high-level general workflow for clinical genetics or genomics laboratories employing next-generation sequencing (NGS) and variant classification is illustrated in **Figure 1**. NGS can

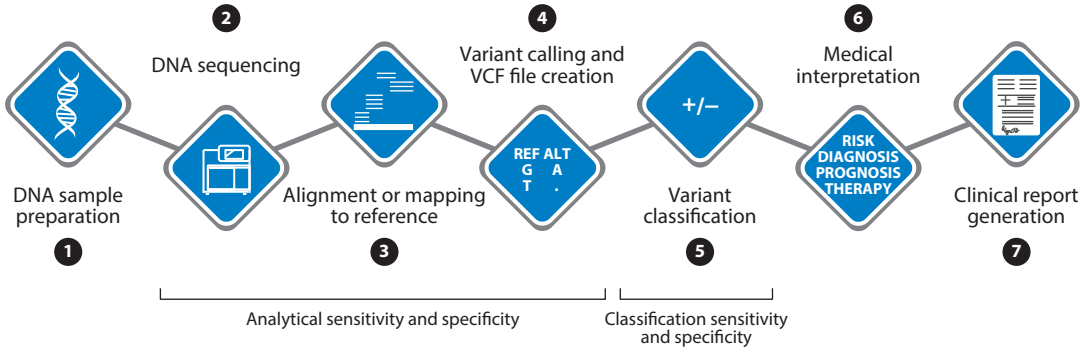

Figure 1

General workflow of clinical genetic or genomic next-generation sequencing tests that require classification of variants. Variant filtering steps can also occur at various points in the workflow. The test characteristics reported by laboratories are almost always limited to gene selection and analytical sensitivity and specificity, though classification sensitivity and specificity directly contribute to test accuracy as well. Abbreviation: VCF, variant call format.

detect variants that are germline or somatic. Germline variants originate from a reproductive cell and are found in the DNA of every cell of the offspring; somatic variants, sometimes called acquired variants, occur after conception and are not found in every cell of the body. Variants are discovered by comparing sequenced DNA to a reference sequence, where any deviation from the reference sequence is considered a variant. A process referred to as aligning or mapping occurs to piece together the small read lengths of sequenced DNA and determine how they correspond with the larger reference genome sequence. Variants are then called—i.e., described with a set of coordinates and other descriptors. Data filters may be used before and after this step, and then variant classification scientists classify each unfiltered variant within the context of the genetic or genomic test application. In the traditional clinical setting, a medical interpretation is then prepared, with oversight by board-certified PhD or MD directors, and a clinical report is generated and signed out for return to the receiving clinician.

In most of the world's clinical genetics laboratories, the current reference sequence for germline variants is the GRCh38 (hg38) reference genome, described by the Genome Reference Consortium. Updated sub-versions are periodically released that may or may not be updated and validated in clinical laboratory workflows. Some clinical laboratories continue to use older Genome Reference Consortium genome builds or their own reference sequence. Approximately two-thirds of the GRCh38 genome build (and of the older GRCh37 build) is derived from a single individual of approximately 50% European and 50% African ancestry (30, 63). Several countries are working to define genomic features that are unique to their populations, or to create their own reference genomes that are better matched to their primary populations (29, 33, 37, 38, 48). An African pangenome has been constructed and contains approximately 10% more DNA than the current GRCh38 reference genome (70).

In the United States, solutions have not yet been broadly implemented in clinical laboratories to account for sequence data in individual genomes that do not align or map to the current reference genome. In NGS, these sequence data are most often discarded when variant call format (VCF) files are generated (see **Figure 1**). Advances are sure to be made, however, as the National Institutes of Health (NIH) National Human Genome Research Institute pushes forward its newer initiative to create a pangenome, which will be a more complete reference sequence to represent 350 genomes from the human population (26). Advances in newer sequencing technologies, particularly long-read technologies, will play a significant role in this effort (46).

The identification, or calling, of both germline and somatic variants is further complicated by persistent limitations in the sequencing and bioinformatic tools that generate VCF files (12, 35, 41). Although NGS often outperforms other methods (53), between 0.1% and 2.6% of detected variants derived from the dominant Illumina NGS platform are incorrect (1, 17, 25, 39, 47, 49, 59, 60). Typical somatic NGS clinical proficiency testing, which evaluates clinical laboratories for the ability to detect the simplest of variants, shows detection sensitivities ranging from 93.9% to 100% (19, 50, 72). Even with identical sequencing reads, a 2020 study comparing common variant calling software found only a 93.8% concordance of single-nucleotide polymorphisms and an 83.5% concordance of insertions and deletions (77).

In addition, different calling pipelines describe some variants differently from other calling pipelines (76). Thus, the downstream variant classification step may be confounded because the variant classification scientist unknowingly does not match relevant data in databases and the literature due to nomenclature differences, with nomenclature being potentially cataloged incorrectly in the reference database or literature and/or by the testing laboratory. Notably, the frequently used Genome Aggregation Database (gnomAD) left-aligns DNA insertions and deletions, in accordance with bioinformatic norms, whereas the Human Genome Variation Society nomenclature right-aligns these same variants, in accordance with clinical norms. Thus, the same insertion or

deletion variant may be named differently using two sets of norms. Even well-known pathogenic variants, such as the cystic fibrosis *CFTR* F508del pathogenic variant, might be missed by bioinformatic pipelines that have not correctly controlled for left- and right-alignment issues (10, 57).

Globally, many laboratory scientists and computer scientists are working to rectify variant detection and variant calling issues that lead to false negative and false positive results. In the meantime, variant classification scientists most often treat variants that queue at the beginning of their workflows as true variants and proceed with laboratory protocols to classify these variants.

3. VARIANT CLASSIFICATIONS ARE TIERED TO REFLECT CERTAINTY

3.1. Germline Variant Classification Tiers

Variant classification has traditionally been treated as "art of medicine" rather than science, which is why it has gone largely unregulated in the United States. Art of medicine falls under clinician licensure rather than Clinical Laboratory Improvement Amendments (CLIA) or College of American Pathologists (CAP) accreditation. To assist in the art of medicine, in 2000 the American College of Medical Genetics and Genomics (ACMG) published an original set of variant classification standards for monogenic disorder germline variants (2), which was followed by a 2007 revision (65). While the general concepts of these standards were adopted by many laboratories, the implementation details and classification terminology, which had not been provided, varied greatly across laboratories.

In 2013, a workgroup consisting of members of the ACMG, CAP, and the Association for Molecular Pathology (AMP) organized to develop standard terminology for germline variant classifications and to make evidence requirements and classification algorithms more stringent. The result was a five-tier classification system. These classifications could be calculated using an additive algorithm from specified evidence categories labeled stand-alone, very strong, strong, moderate, or supporting. The 2015 guidelines were titled "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" (64). The five tiers of germline variant classifications were pathogenic, likely pathogenic, benign, likely benign, and (variant of) uncertain significance (VUS or VoUS). The guidelines recommended that variants be reported with respect to the relevant genetic disorder and inheritance pattern. It was also stressed that professional judgment should be used when implementing and adapting the guidelines.

A survey of the germline clinical genetics community, as described in the 2015 guidelines, determined that a 90% certainty was sufficient to describe a variant as either likely pathogenic or likely benign (64). Reference was also made to the International Agency for Research on Cancer's recommendation that a 95% level of certainty be applied to variants classified as pathogenic (61). However, sufficiently large orthogonally validated truth sets are not publicly available to actually quantify laboratories' certainty of variant classifications; therefore, the certainty concept remains aspirationally qualitative.

In most clinical scenarios, both the pathogenic and likely pathogenic variant classifications are considered positive results that may alter care, with appropriate consideration to inheritance patterns; the benign, likely benign, and VUS classifications are generally considered results that do not change care. Certain clinical scenarios may warrant the treatment of a VUS as a candidate positive finding when other diagnostic opportunities have been exhausted, such as in the pediatric setting for patients in rapid decline.

Copy number variants—i.e., large deletions and gains (such as insertions, duplications, and triplications) at the exon and gene size scale—have long been detected by technologies other than

sequencing technologies. However, NGS has advanced to being able to detect some of these types of larger variants. Until 2019, copy number variants had their own classification categories. However, in the 2020 publication "Technical Standards for the Interpretation and Reporting of Constitutional Copy Number Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen)" (66), the classification system was uncoupled from individual patients and aligned to the five-tier classification system, matching the 2015 ACMG/AMP sequence variant classification guidelines (64).

3.2. Somatic Variant Classification Tiers

Somatic variants, also known as acquired variants, may be used in the cancer setting for diagnosis, for prognosis, and/or to guide therapy selection. Detected variants in these tests are frequently reported along with variant allele frequencies measured within each unique patient sample to estimate the probability of variants being germline, and to evaluate tumor clonal diversity. Sometimes "normal" patient matched tissue is also used as a comparison, with the aim to distinguish somatic variants from germline variants. However, what is considered normal tissue is not straightforward, and different laboratories may approach this challenge differently, yielding different results, while some laboratories opt to use computational methods instead (21).

The first major somatic variant classification guidelines were published in 2017 and remain current, though an update effort began organizing in 2020. The 2017 standards have not been broadly adopted by laboratories (7, 71), though the general concepts are found in almost all somatic testing laboratories. Already established laboratories with optimized protocols have been less likely to adopt the guidelines. The 2017 somatic classification guidelines, titled "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" (44), recommends a four-tier system with the following evidence requirements (as summarized in figure 2 in those guidelines):

- Tier I: variants of strong clinical significance
  - Level A evidence:
    - FDA-approved therapy
    - Included in professional guidelines
  - Level B evidence:
    - Well-powered studies with consensus from experts in the field
- Tier II: variants of potential clinical significance
  - Level C evidence:
    - FDA-approved therapies for different tumor types of investigational therapies
    - Multiple small published studies with some consensus
  - Level D evidence:
    - Preclinical trials or a few case reports without consensus
- Tier III: variants of unknown clinical significance
  - Not observed at a significant allele frequency in the general or specific subpopulation databases, or pan-cancer or tumor-specific variant databases
  - No convincing published evidence of cancer association
- Tier IV: benign or likely benign variants
  - Observed at significant allele frequency in the general or specific subpopulation databases
  - No existing published evidence of cancer association

The same four-tier system was used in the 2019 guidelines titled "Technical Laboratory Standards for Interpretation and Reporting of Acquired Copy-Number Abnormalities and Copy-Neutral Loss of Heterozygosity in Neoplastic Disorders: A Joint Consensus Recommendation from the American College of Medical Genetics and Genomics (ACMG) and the Cancer Genomics Consortium (CGC)" (52). Clinicians feel most confident in disclosing both tier I and tier II results to patients to inform patient care (24).

4. DISCORDANCE AND POSITIVE BIAS IN GERMLINE AND SOMATIC VARIANT CLASSIFICATION

4.1. ClinVar Allows Cross-Comparison of Classifications

In 2013, the NIH's National Center for Biotechnology Information expanded the concept of locus-specific variant databases (73, 74) and launched ClinVar as a public archive for clinical laboratories, researchers, expert panels, and others to archive their classifications of variants across all genes in the germline context (42) and more recently in the somatic context. Sometimes ClinVar participants provide minimal or detailed evidence summaries with their submissions. A rating system has developed to reflect concordance and/or confidence in classifications, and the Food and Drug Administration (FDA) now approves certain classifications found in ClinVar that are made by FDA-sanctioned expert committees (13). These gene- or disease-specific committees vary in their reliance on evidence as opposed to expert opinion. ClinVar facilitates direct comparison of discordance among laboratories that submit to the archive.

4.2. Discordance and False Positives in Germline Classifications

The 2015 ACMG/AMP germline variant classification guidelines provided additive algorithms and criteria for classifying (likely) pathogenic variants and (likely) benign variants (64). Variants unable to be classified with these criteria or with unresolvable evidence conflicts remain VUSs until they are eventually clarified. Some evidence criteria carry recommendations for quantitative or semiquantitative thresholds. More commonly, however, the evidence criteria are conceptual in nature. This is an inescapable result of clinical variant classification science being developed in the absence of large truth sets. In the absence of detailed and orthogonally validated truth sets, expert opinion is used to fill knowledge gaps. Unsurprisingly, studies comparing the application of the 2015 ACMG/AMP classification guidelines across different laboratories reveal discordance.

Of particular importance is discordance that has a clinical impact, such as discrepancies between pathogenic/likely pathogenic variants and VUSs. These discordances occur in a range of 1.5% to 10.8% among clinical laboratories (5, 6, 18, 45). Examination of trends in discordance over time shows that since the development and publication of the 2015 classification guidelines, overall clinical discordance has decreased to around 2.8% (75). This may also be a product of more communication between laboratories and integrated use of existing ClinVar classifications (32). Note, however, that concordance does not guarantee accuracy. In a 2018 study, researchers demonstrated that approximately 11.5% of genetic disorders already had enough variants cataloged as (likely) pathogenic in ClinVar to exceed the population prevalence of genetic disorders (69). Considering that ClinVar is an incomplete archive of genetic variants, the study's estimation of impact from false positive variant classifications represents a lower bound. The study also concluded that the false positive classification burden arose mostly from misclassified rare variants rather than misclassified common variants.

4.3. Discordance and False Positives in Somatic Classifications

In tumor genomic profiling tests, also known as somatic tests, multiple studies have shown extremely high discordance rates among different laboratories (16, 27). When comparing final test results, it is difficult to determine which of the multiple complex laboratory steps lead to discordance, though we know that current somatic variant classification technologies suffer from many classification false positives where random acquired variants are often overinterpreted as driver variants (43). In one of the largest comparative studies, published in 2019, three mimetic non-small-cell lung cancer samples with six variants were sent to 101 different laboratories. Two variants had above 85% concordant classifications, while four were highly discordant. Only one-third of the laboratories returned results correctly identifying all six expected variants (27).

5. EFFORTS TO IMPROVE VARIANT CLASSIFICATION METHODOLOGY

5.1. Public Efforts

The Clinical Genome Resource (ClinGen) consortium has been actively working to develop publicly available tools, procedures, and working groups to advance the field of genetic variant classification in both germline and somatic settings (14). Other disease specialty consortia, such as the Evidence-Based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium (**https://enigmaconsortium.org**) in hereditary cancer, have similar goals, as do many independent clinical researchers (e.g., **https://cser-consortium.org**). The ACMG, CAP, AMP, and other societies are working to update their general and disease-specific variant classification guidelines in both germline and somatic settings. These groups have varying levels of access to quality-controlled data, resources, and statistical support. Unfortunately, in the absence of large orthogonally validated truth sets, with data rich at the patient and sequence quality levels, and without sufficient resources to leverage these data, even the excellent work of these consortia and clinical scientists is often informed more by expert opinion than by sufficiently powered evidence. In time, as more valid evidence is compiled and studied, variant methodology will improve and classifications will become more accurate. For example, promising early results have been emerging from researchers working with data compiled within the UK Biobank (68).

5.2. Private Efforts

Several commercial clinical genetics laboratories with large, quality-controlled databases of paired sequence data and patient phenotype and family history data actively work to improve the accuracy of variant classification. The internal databases of these laboratories allow internal researchers and outside collaborators to develop advanced variant classification techniques, which these laboratories often present or publish for product promotion and peer review. These classification and reclassification methodologies will improve over time, but currently they often represent the most advanced variant classification methodologies in clinical genetics and genomics. When carefully considered, these commercial methodologies should inform the drafting of public guidelines.

Particularly important advances in germline variant classification have been developed and made public by several high-volume commercial laboratories. Methodologies of note (with their brand names in parentheses) include Myriad Genetics' mutation co-occurrence (M-CO) analysis and its history-weighting algorithm (Pheno) for the reclassification of variants in high-penetrance autosomal dominant disorders (20, 22, 54, 62). Invitae's comprehensive refinement of the ACMG/AMP variant classification criteria (Sherloc) is an exemplary iteration of the

guidelines (58). Myriad Genetics' (unbranded) (11) and Ambry Genetics' (+RNAinsight) (36) pairing of germline DNA and mRNA sequencing to disambiguate RNA-splicing variant classifications is helping to reclassify many variants. In the oncology somatic space, Foundation Medicine's patient-level database of variants and variant allele frequencies (FoundationCore), in combination with Flatiron Health's patient-level clinical outcomes database, is accelerating discovery for cancer medicines and biomarkers (3). Other laboratories are also making significant advances. While the development of private commercial databases and methodologies continues to be a contentious topic among some stakeholders, commercial laboratories committed to publishing their work are undeniably delivering important advances to the field of variant classification.

6. KEY LESSONS FROM ESTIMATES OF POSITIVE PREDICTIVE VALUE IN GERMLINE NEXT-GENERATION SEQUENCING TESTS

6.1. Modeling Positive Predictive Value to Include Both Analytical and Variant Classification Test Characteristics

Sensitivity and specificity are familiar terms when describing diagnostic test characteristics. Mathematically combining these test characteristics with the true genetic disorder prevalence in the population being tested provides the clinical positive predictive value (PPV) of the diagnostic test, as defined with the standard equations shown in **Figure 2*a***. In the modeling demonstrated here, clinical PPV is the percentage chance that a patient's variant, when reported by the laboratory as positive, is truly positive. Positive variants are those classified as pathogenic or likely pathogenic.

At present, most genetic and genomic sequencing test reports that have incorporated variant classification are accompanied by technical specifications that describe the test coverage and technical limitations of the test, including metrics of analytical sensitivity and specificity. These analytical metrics relate to the performance of the sequencing instrument and the variant calling pipeline (**Figure 1**). Test specifications rarely, if ever, include metrics of novel variant classification sensitivity or specificity since variant classification lacks the public truth sets needed to generate classification sensitivity and specificity metrics. Therefore, in the absence of variant classification sensitivity and specificity estimates, the clinical PPVs of most modern genetic and genomic NGS tests have been left unexplored until now.

Two of the authors (H.H.G. and J.M.E.) created a framework to model clinical PPV in NGS germline sequencing tests that incorporates both the reported analytical performance of the tests and informed estimates of variant classification sensitivity and specificity. The framework of inputs and outputs is diagrammed in **Figure 2*b***. For simplification, the five-tier germline variant classification system was converted into a binary system based on the most common clinical scenario, where variants classified as pathogenic and likely pathogenic are likely to alter clinical care (positive) and variants classified as likely benign, benign, or VUSs are not likely to alter clinical care (negative).

In the simplified model, a germline genetic disorder is caused exclusively by autosomal dominant simulated genes and is entirely detectable within a simulated 100-gene diagnostic panel. The hypothetical gene sequencing lengths are evenly distributed between 14.5 kb and 1.1 kb, modeled after the diagnostic sizes of the large *ATM* gene and the small *CDKN2A* gene, respectively. A hypothetical gene is randomly selected to capture 30% of the genetic disorder and represents the single-gene test scenario. The expanded 20-gene panel with additional randomly selected genes captures 63% of the genetic disorder, and the 100-gene panel captures 100% of the disorder.

The gnomAD data for the large *ATM* gene (9) is used to set an average rate of variation of 0.32 variants per kilobase in the model, with each allele contributing 0.16 variants per kilobase. While variation is not evenly distributed across the genome, this assumed constant rate of variation

a

$$\text{sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \qquad \text{specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

$$\text{positive predictive value} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{positive predictive value} = \frac{\text{sensitivity} \times \text{prevalence}}{[\text{sensitivity} \times \text{prevalence}] + [(1 - \text{specificity}) \times (1 - \text{prevalence})]}$$

b

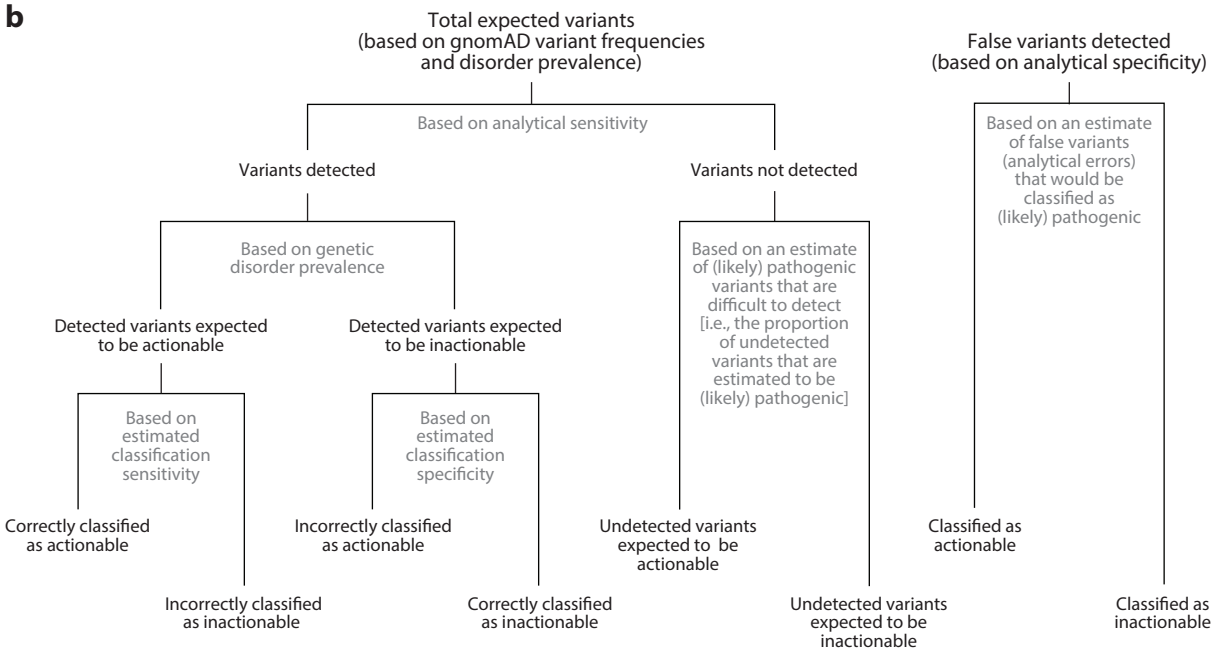

Figure 2

A framework to calculate the positive predictive value of germline autosomal dominant gene sequencing tests that include variant classification. Panel *a* shows the definitions for sensitivity, specificity, and positive predictive value, where prevalence means the true genetic disorder prevalence in the testing population. In panel *b*, for the analytical (detection) portion of the test, a positive variant is any detected variant, regardless of classification. For the variant classification portion of the test, a positive variant, which may alter care, is likely pathogenic or pathogenic; a negative variant, which is not likely to alter care, is likely benign, benign, or a variant of uncertain significance. The model uses variant allele frequencies and genetic disorder prevalence, along with reported and estimated sensitivities and specificities, to frame the various outcomes relevant to genetic tests. Abbreviation: gnomAD, Genome Aggregation Database.

serves as an acceptable basis for a simplified illustrative model. The simulated autosomal dominant genetic disorder is set as having a general population prevalence of 1%.

The analytical sensitivity/specificity of the simulated 20-gene panel for high-, medium-, and low-stringency laboratories is set to 99.99%/99.99%, 99.95%/99.95%, and 99.00%/99.00%, respectively, to match laboratory-reported test characteristic ranges for real clinical laboratory sequencing tests of somewhat similar size (5, 15, 28, 55, 56). These analytical test characteristics are then extrapolated to the single-gene and the 100-gene modeled tests.

The rate of novel variant classification true positives, false positives, true negatives, and false negatives is estimated from published counts at the patient level of initial variant classifications, and then reclassifications in hereditary cancer from a major laboratory over a 10-year period (51). For example, pathogenic/likely pathogenic classifications that were later downgraded to VUS (a positive-to-negative reclassification in our model) are categorized as false positives, while pathogenic/likely pathogenic classifications that were not reclassified during the 10-year study are considered true positives. Although reclassification counts in this 10-year study cohort will increase over time, the data come from a laboratory that has consistently operated a proactive variant reclassification program. The laboratory's published classification and reclassification counts therefore serve here to optimistically estimate the classification sensitivity and specificity of a very stringent genetic test, according to the standard equations illustrated in **Figure 2*a***. Such an analysis of these data results in a novel variant classification sensitivity/specificity of 98.5%/99.995%. Based on these estimates, we model the variant classification sensitivity/specificity for high-, medium-, and low-stringency laboratories at 98.00%/99.99%, 99.00%/99.98%, and 100.00%/99.97%, respectively. Note that higher-stringency laboratories gain specificity in part by decreasing sensitivity, which in practice typically means requiring higher levels of evidence for variants to be classified as (likely) pathogenic instead of as VUSs, as compared with lower-stringency laboratories.

Variant classification sensitivity and specificity, as treated here, are at a per-observation level and not a per-unique-variant level. Each variant observation is treated independently. For example, if there are two unique variants, variant A and variant B; variant A is a false positive, while variant B is a true positive; and in 10 patients, variant A is observed 1 time and variant B is observed 9 times, then 1 out of 10 variant classifications is a false positive, and 9 are true positives. This patient-centric approach is why the classification specificity estimates used here appear very high to variant classification scientists accustomed to wrestling with the evidence of only rare variants. The cumulative allelic frequencies of the variants that are very common and therefore very easy to accurately classify as benign contribute a high observation count of true negative variants to the specificity equation (**Figure 2*a***). The cumulative allelic frequency of a few common variants often exceeds the cumulative allelic frequency of many rare variants by more than an order of magnitude, as exemplified when using an arbitrary but realistic threshold in the *ATM* gene (**Figure 3**).

6.2. Key Takeaways from Positive Predictive Value Modeling

The autosomal dominant clinical PPV curves for the single-gene, 20-gene, and 100-gene simulated test scenarios are shown in **Figure 4*c***; the contributing effects of the analytical and classification sensitivities and specificities are illustrated in **Figure 4*a*** and **Figure 4*b***, respectively, and

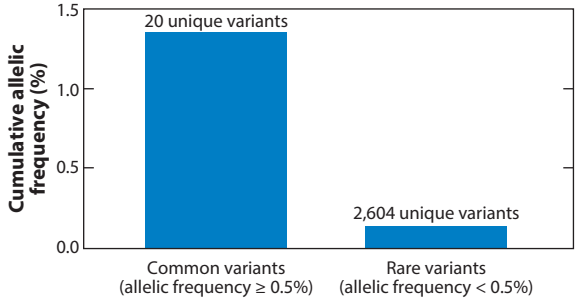

Figure 3

Cumulative allelic frequencies of common and rare variants in *ATM* (9).

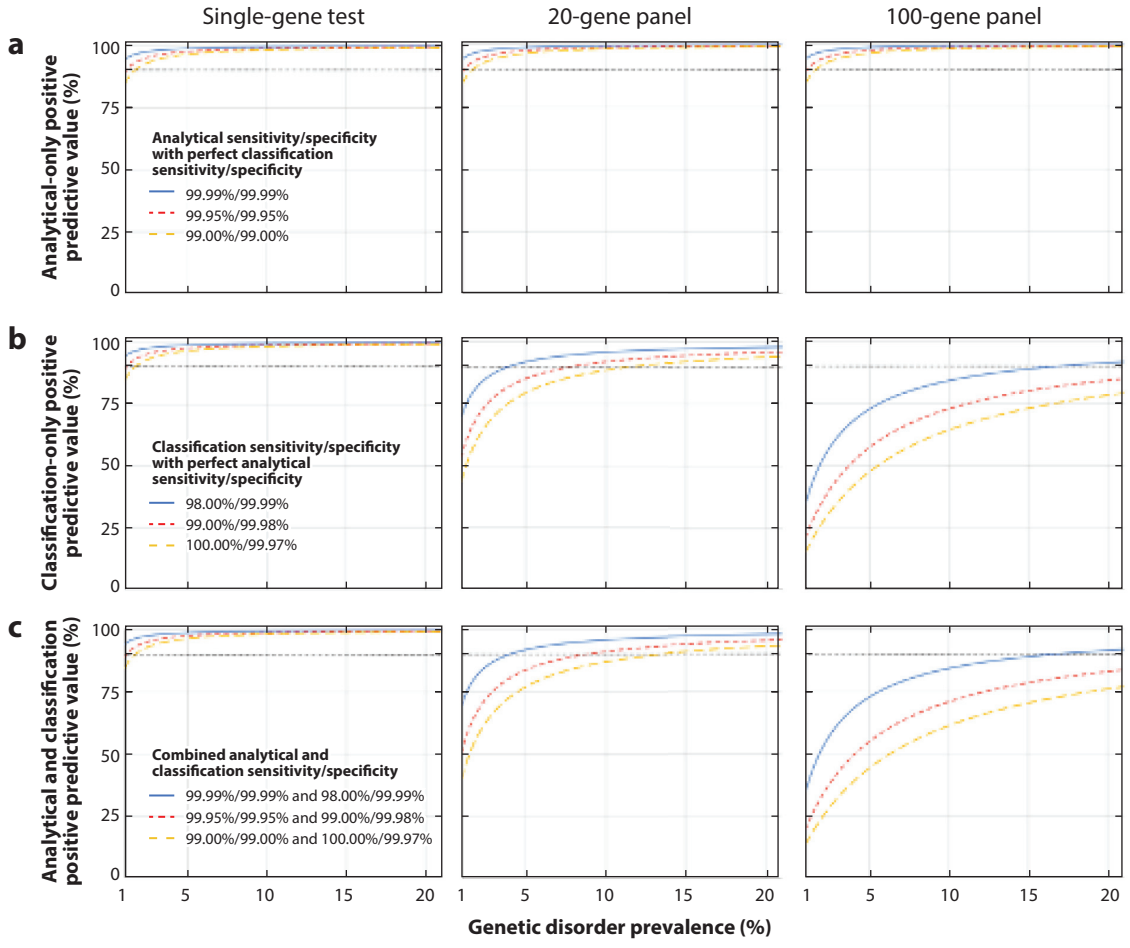

Figure 4

Modeled clinical positive predictive value curves for autosomal dominant next-generation sequencing panels of varying sizes using simulated data. Key assumptions and inputs are described in the main text. In each panel, the horizontal dotted black line corresponds to a positive predictive value of 90%, which genetics professionals have voted to be an acceptable level of confidence in the truthfulness of a positive test result (64). (*a*) The impact of analytical imperfections on variant detection, where the classification sensitivity and specificity are both set to 100%. The three colored lines are derived from the indicated analytical sensitivities/specificities in the 20-gene panel and then extrapolated to the two other tests. (*b*) The impact of variant classification imperfections on accurately reporting a positive variant (likely pathogenic or pathogenic), where the analytical sensitivity and specificity are both set to 100%. As in panel *a*, the three colored lines are derived from the indicated classification sensitivities/specificities in the 20-gene panel and then extrapolated to the two other tests. (*c*) A realistic scenario that incorporates both analytical and variant classification imperfections, as combined from panels *a* and *b*. The solid blue line represents a high-stringency laboratory, the dashed red line represents a medium-stringency laboratory, and the dashed yellow line represents a low-stringency laboratory in our model.

Figure 4*c* shows the combined model, including estimates that represent the PPVs from high-, medium-, and low-stringency laboratories. Three key considerations for germline NGS sequencing tests emerge.

6.2.1. Stringent variant classification is critical to minimize false positives. In comparing **Figure 4*a*** with ***b***, we see that the most significant contributor to false positive test results is imperfect variant classification specificity, as compared with analytical sensitivity and specificity or

classification sensitivity. Therefore, the common practice of judging NGS test performance only on analytical test characteristics and gene selection is dangerously incomplete. Minor imperfections in variant classification specificity produce drastically more false positive test reports than typical imperfection ranges in analytical performance. (While not shown, large decreases in variant classification sensitivity increase false negative reports. However, as described above, current variant classification methodologies bias toward positives rather than negatives.)

While this key point was made by modeling germline panels, a similar point about PPV and variant classification specificity can be made in the somatic context. We hope that future updates in both germline and somatic variant classification guidelines will assist in increasing test PPV. Section 7 addresses specific items that are currently lowering the specificity of variant classification in many laboratories.

6.2.2. Positive predictive value drops significantly when clinical tests are used to screen the general population. Comparison of all panels in **Figure 4** demonstrates that PPV decreases rapidly when genetic tests are moved from a clinical application to a general population application, with variant classification false positives being the primary contributor. When the prevalence of a genetic disorder in a testing population is lower than a test's false positive rate, more false positive than true positive results will occur. In diagnostics, this is known as the false positive paradox, and it is often counterintuitive to clinicians who are accustomed to using genetic tests in a clinical setting, where the testing population is significantly enriched for the expected genetic disorder. In our model, the false positive paradox threshold is readily crossed in both the 20-gene and 100-gene panel scenarios as the genetic disorder prevalence begins to approach the modeled population prevalence of 1% (**Figure 4*c***). Stakeholders must consider the inevitable drop in PPV when considering implementation of diagnostic tests as general population screening tools (31). Similarly, policy makers who are decreasing a priori risk requirements for genetic testing candidates must carefully weigh the consequences of the resulting decrease in testing PPV. To adapt, laboratories can increase their variant classification specificity to account for the lower-risk population and reevaluate previous classifications in their databases with higher stringency.

6.2.3. Larger gene panels increase risk for false positive reports. Comparing the three panels in **Figure 4*c*** reveals that increasing gene panel sizes amplifies the ratio of patients receiving false positive results to patients receiving true positive results, primarily due to imperfect variant classification. Akin to Russian roulette being repeatedly played, the risk of a false positive test result grows with each gene added to a test panel. Celebrated increases in the diagnostic yield of many newer supersized sequencing tests may be due in part to false positive variant classifications. Therefore, a careful risk/benefit analysis must be considered when increasing the size of gene panels or moving to whole-exome or whole-genome tests for patients. Test developers and test users must determine whether the gain in test sensitivity from adding genes is worth the risk of increasing the frequency of false positive results. The right balance and choice of genes on a panel will depend greatly on the downstream clinical implication of false negatives and false positives for each candidate gene. Also, consideration should be given to the availability and accuracy of orthogonal tests that could be used to verify the findings of a gene panel.

Table 1 models a clinical setting where patient phenotypes and/or personal and family history enrich the testing population to an a priori 10% genetic disorder prevalence. The table includes two rows that give the average number of patients tested for each true positive or false positive result obtained, respectively. For example, clinicians deploying a single-gene test from the most stringent laboratory will see 34 patients for each true positive test result, but if they switch to the 100-gene panel from the same laboratory, they will then see 10 patients for each true positive

Table 1 Frequency of patients receiving true and false positive variants from the simulated next-generation sequencing panels in a clinic population with a true 10% genetic disorder prevalence

	Single-gene test			20-gene panel			100-gene panel		
	High	Medium	Low	High	Medium	Low	High	Medium	Low
Number of common variants	1.75	1.75	1.75	30.86	30.87	30.87	167.85	167.87	167.90
Number of rare variants	0.17	0.17	0.17	2.97	2.97	2.97	16.16	16.17	16.17
Total number of variants	1.91	1.91	1.91	33.84	33.84	33.85	184.01	184.04	184.07
Number of patients per reported true positive	34	34	34	13	13	13	10	10	10
Number of patients per reported false positive	5,285	2,399	1,554	295	134	87	54	25	16
Number of rare variants classified per false positive	889	403	261	877	398	258	875	397	257
Number of common variants classified per false positive	9,227	4,189	2,715	9,103	4,134	2,679	9,087	4,126	2,674

In the column headings, high, medium, and low denote high-, medium-, and low-stringency laboratories; the test characteristics of these laboratories are defined in **Figure 4**. Counts are given according to the number of variant observations in patients rather than unique variant counts.

test result. This increase in diagnostic yield seems ideal. However, when switching from using the single-gene test to the 100-gene panel, the clinicians go from returning false positive test results to a patient very infrequently (1 in 5,285 tested patients) to returning false positive results somewhat regularly (1 in 54 tested patients). Unfortunately, to most clinicians, the true positives are indistinguishable from the false positives. Trade-offs must be carefully considered when choosing between smaller and larger gene panels, even when using highly stringent laboratories.

7. EXPERT OPINIONS ON THE CURRENT STATE OF THE FIELD

Given that variant classification guidelines rely heavily on expert opinion, it is fair to consider expert opinion when exploring the strengths and limitations of the guidelines. The expert opinions shared below are informed mainly by germline testing, but many of the comments also relate to somatic settings.

Elaine Lyon (E.L.) and Madhuri R. Hegde (M.R.H.) were among the coauthors of both the 2015 ACMG/AMP and 2007 ACMG guidelines for germline sequence variant classification (64, 65). They remain active leaders in the global clinical genetics community and continue to develop protocols in their own clinical laboratories. For additional perspective on current practices, Christine M. Stanley (C.M.S.), Iain D. Kerr (I.D.K.), and Julie M. Eggington (J.M.E.) comment. C.M.S. is a board-certified laboratory director, and I.D.K. is a structural biologist. C.M.S. and I.D.K. are both leaders in high-throughput commercial variant classification, and J.M.E. has extensive experience in quality checking and consulting on germline and somatic clinical variant classification pipelines, primarily in the high-throughput commercial laboratory space.

The commentary provided by the authors is their own; some authors may not agree with the comments of other authors, and the comments do not necessarily reflect the opinions of the authors' employers or the organizations, societies, or committees of which they are a part. The authors also reserve the right to change their opinions.

7.1. The 2015 Germline Variant Classification Standards and Guidelines from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

7.1.1. Historical context for the creation of guidelines. M.R.H.: A major driving force for creating [the 2015] guidelines for variant classification was reductions in the cost of sequencing. Subsequent to this affordability, we began to encounter a rise in population data, necessitating rigor in how we approached variant classification. We really needed to look at all the variants in these population data, because many of them would be benign.

E.L.: The variant classification landscape prior to the 2015 publication of the sequence variant guidelines, truthfully, did not change too much. Previously, we lacked a detailed standard operating procedure (SOP). We basically took into consideration different ideas, and some differences in the strength of the evidence and how to apply it occurred. One of the surprising aspects of composing the 2015 guidelines was realizing that across laboratories, we had orthogonally come to similar conclusions. Creating the 2015 guidelines was a process, and some concepts we could not address completely, but we could at least place those concepts in a framework where other members of the broader variant classification community could address them. Though some confusion persists on how to use the guidelines, I am happy with how they turned out as an important first step. In a way, the guidelines started the science of variant classification and interpretation.

7.1.2. Major successes of the guidelines. M.R.H.: The uniform acceptability of the guidelines was wonderful to see because the committee spent a lot of time creating them, including two years of debate and dedicated thoroughness. Additionally, I am excited that the community wants to debate the guidelines. It is critical for the progress of the field that people take pieces of the guidelines and debate them. I receive global feedback from colleagues both agreeing and disagreeing with aspects of the guidelines. I have seen that clinical decisions can drive the interpretation of a VUS, which can then engender its own debate, lack of endorsement from some for such decisions, and rebuttal, which is essential for progress.

C.M.S.: I most appreciated the premise of bringing initial structure and standardization to the classification of variants at a time when NGS was ramping up heavily and therefore the number of variants for classification was exploding. Without proper guidance and established expectations in the field, we might have set genetic testing back years, because if labs were arriving at different conclusions, then the utility of genetic testing as a whole may have been seriously questioned. The straightforward definitions of the lines of evidence used in variant classification coupled with good explanations and examples served the industry well in the beginning to allow for adaptation, rather than an overburdensome and overly complicated system, which may have delayed adoption. I preferred the five-tier system rather than more complicated proposals and the option to weight the lines of evidence up or down to prevent the under- or overclassification of variants.

7.1.3. Implementation challenges of the guidelines. E.L.: After the 2015 guidelines came out and were implemented, we were better able to see which components elicited more debate and, in some cases, more confusion. This indicated areas where we needed to provide improved clarity. Since the committee that created the 2015 guidelines was so experienced in variant classification, perhaps we did not anticipate the questions some of our less experienced colleagues had.

M.R.H.: A criticism from the overall community that we sometimes receive is that the [2015] guidelines do not apply to one particular gene or another, even though the guidelines are intended to be more generic. Specific diseases, genes, variants, and heterogeneity need to be considered individually. I am uncertain if creating a guideline for every gene—which is happening right now—is

going to be useful for clinical labs. Software and artificial intelligence will be necessary to support this, and so will the application of one's skill set to variant interpretation. Potential future overreliance on software concerns me. Moving forward, clinical lab directors need to play an important role, especially as in my experience on the CAP committee and in NIH proficiency testing, most errors arise from labs using automated software to classify variants. The software cannot do it all. Also, when evaluating evidence in publications, the whole paper needs to be taken into consideration. Piecemealing ought not have a role in these assessments, and I would argue that it can even be dangerous to segregate everything into a million pieces.

C.M.S.: Some of the challenges we faced with adoption was a lack of clarification of how to find the data for the recommended lines of evidence and how to utilize some of the lines of evidence; for example, co-occurrence evidence is very challenging. In practice, this led to using a subset of the available lines of evidence. Additionally, we needed a bit more guidance on using the scoring algorithm for cases when there was contradiction; for example, when there were two lines of evidence for benign classification, guidelines indicate the variant as benign even when there is conflicting evidence. Also, there was some confusion on selecting lines of evidence of the same type, which amounts to double-dipping and overclassification. We also were challenged with transcript selection. When reviewing variants, it was tempting to be overly cautious and select the longest transcript or the one with the highest variant effect predictor score. This resulted in some cases of overscoring a variant as pathogenic by using a transcript that was not one that was of biological significance. However, overall this was a great first draft, and there was always a plan to update this first draft after the genetics community worked with it.

7.1.4. Application of the guidelines beyond the United States. M.R.H.: We presented the guidelines on variant classification internationally and have discovered that different countries take very different approaches to the guidelines. For example, in India, they think these guidelines are too stringent, perhaps because many clinicians practice medicine based on their skill set and are not reliant on software because it is not available. In this context, understanding the interpretation of pathogenicity can mean proceeding with treatment, as there is no mechanism to provide feedback through the guidelines. Furthermore, the practice of genetics in Europe is quite distinct from the practice of genetics in the United States. Other than one or two companies, whole-genome sequencing has not been adopted. The debates that arise from this global perspective are critical. Sometimes, in the United States, we tend to narrow our focus to how we practice medicine. We have an opportunity to evaluate these guidelines in new contexts when we are open to global feedback.

7.2. ClinVar

7.2.1. Achievements. M.R.H.: I have been involved in ClinVar from early on and am thrilled the submission rate now is quite high. ClinVar has come a long way with regard to how laboratories are now understanding what to submit. Additionally, laboratories seem to understand the difference between ClinVar and HGMD (the Human Gene Mutation Database).

C.M.S.: ClinVar has contributed to transparency in the industry.

E.L.: ClinVar is becoming a pretty good database when multiple people are in accordance on whether a variant is benign. When there are no contradictions across entries, ClinVar is a particularly efficient database. In ClinVar, I really appreciate when people who have submitted have provided references with their description, enabling me to go to the original source. This improves transparency for the rationale, rather than assuming I will just take someone's word for their description.

7.2.2. Challenges and outdated classifications. M.R.H.: ClinVar still needs significant cleanup. For example, there are still variants that we know are benign that remain classified as pathogenic on ClinVar. Given the global use of ClinVar, nuances persist that can be dangerous.

C.M.S.: The biggest failures in ClinVar are (*a*) tagging variants added to ClinVar as pathogenic based on the variant being discussed in an OMIM review, rather than tagging them as "informative"; (*b*) excessively pressuring labs to deposit all or a large percentage of variants, rather than letting labs choose the variants that they thoroughly reviewed; and (*c*) allowing labs to continue to add variant classifications to ClinVar when they have had a large percentage of incorrect classifications in the past. I would also recommend an ability to provide feedback on agreement or disagreement with another lab's review and allow labs to contest a review or recommend redacting a review. Inaccurate legacy classifications are a big problem for databases such as ClinVar.

J.M.E.: Some clinical labs use ClinVar classifications in their reports without any quality checks, even single observations, then submit those same classifications back into ClinVar as their own. This results in false consensus. Mechanisms should be put in place to stop this practice since an increasing number of labs use consensus as a trigger to pass a variant classification on to clinical reports without review.

I.D.K.: As a community, we should have the decency to say that, given emerging evidence, we wrongly classified a variant. At the time, the classification may have been completely reasonable. Biology is dynamic, and dogma is often challenged (67). I consider incorrect classifications to be in the vast minority of cases; however, we should actively reassess these variants as new data become available.

E.L.: We need to be aware of the difference between contradicting evidence and outdated evidence. If data from 1995 indicate pathogenicity but current data indicate no pathogenicity, the data from 1995 can be considered outdated. This distinction, however, can be confusing.

7.2.3. Considerations for Food and Drug Administration–approved variant classifications. E.L.: It is critical we remember that FDA-approved variant classifications in ClinVar are still expert opinions. Certainly I feel better that a group of experts in a certain disease have reviewed the variants, especially if this comes with a publication describing their approach. My concern, however, is that some people will take FDA approvals as the truth, which they may or may not be.

M.R.H.: It is dangerous to go in the direction of relying on expert committees for FDA-approved classifications in ClinVar, especially for the way approvals are made with points and a star classification. Compounding the issue, publications are occurring for one modification or for modification of a certain criteria, the application of which is uncertain. The potential for someone to then go through the ClinVar SOP program and rely on the star classification without applying critical analysis could create a dangerous situation because a lot more sequencing from diverse populations needs to occur. Though we do not need to worry about variants that are obviously benign, we are additionally moving in the direction of very rare variant burden and polygenic risk scores where the effects of these variants are unknown. We are narrowing our focus too much rather than evaluating a global perspective of what it means to do clinical genetics and clinical testing.

7.3. Improving Variant Classification Science

7.3.1. Priorities for improvement. M.R.H.: For germline, the top priority for updating the guidelines should be to integrate both the copy number variation guidelines and the ACMG sequence variant guidelines. We need to integrate biochemical cytogenetics and molecular genetics into the guidelines. For germline and somatic settings, a priority would be to utilize data on

clinical responses to help establish whether a variant is pathogenic. There should be a feedback loop on the success or failure of precision therapies for individual patients. Third, we need a proper infrastructure for functional genomics. Additionally, in germline population-based analysis, we need to be watchful around benign variants, rare variant burden, and the evolution of polygenic risk scores. Also, appropriate training is a grave unmet need in clinical genetics and genomics, as current training programs in molecular and cytogenetics that last just two years do not prepare candidates to contribute meaningfully in the clinic.

I.D.K.: For loss-of-function genes in both germline and somatic settings, it is important to know the last known pathogenic variant that is supported well with evidence. That helps us understand how much can be deleted at the end of a gene before a pathogenic event occurs. Also, the use of in silico algorithms as a potentially deciding factor in classification has to stop (40). We need, and patients deserve, meaningful functional assays that can predict loss of function and lead to better medical management decisions. Also, about ClinGen—ClinGen is, largely, an academic effort. The context of a high-throughput lab, in biotech, is not apparent to them. I applaud their efforts to reach some form of standardization, but they need to reach out to a wider community to truly effect some kind of change. Working successfully between the ClinGen committees and biotech can be unnecessarily challenging, and wider clinical genetics/genomics engagement could support overcoming these challenges.

C.M.S.: We need a consensus on the correct transcript. Locus Reference Genomic is meant to be that database, but it only contains a fraction of genes. All variant curations should be based on the level of a variant by transcript by disease. Additionally, in germline specifically, we need a consensus on the incidence of diseases associated with genes, which would enable us to determine whether a variant's frequency is too high to be pathogenic.

E.L.: Understanding how reduced penetrance affects germline variant classification is important. What do we do with variants that partially disrupt but still retain some function? We can call those variants "reduced penetrance" variants or variants with "different expressivities." I have used the term "mild variants." But reduced penetrance is one thing the 2015 guidelines did not address. With reduced—or mild—penetrance, we need to think carefully about when such variants should go on a report. I would not call such variant classification "inaccurate," but rather "incomplete," because we simply do not know the potential effects. A lot of variants do something, but at what point do they cause disease?

J.M.E.: Laboratories decrease their variant classification specificity when they use insufficient and/or low-quality data inputs for variant classification (69), and/or use overly permissive variant classification algorithms. Germline variant classification used to rely heavily on limited and unreplicated research databases, publications, and/or case reports. In time, these data sources proved to be unreliable (34). Many of these lines of evidence were demoted in the 2015 ACMG/AMP germline variant guidelines, and are likely to be further demoted in future iterations. Therefore the practice in germline of adding up weak lines of evidence to mimic a stronger line of evidence is inadvisable. In a concerning repeat of history, the 2017 AMP/ASCO [American Society of Clinical Oncology]/CAP somatic variant classification guidelines adopted weak lines of evidence within its tier II category, likely in an effort to increase sensitivity, yet at the expense of increasing false positives. Clinicians use tier II results to treat patients. However, the toll on patients of unnecessary treatment and the increasing costs of targeted oncology therapies provide strong incentives for decreasing false positives. Also, many laboratories overinterpret functional assay data in the literature, which leads to false positives. The decrease in activity or performance of a mutated protein or pre-mRNA splice site as measured by a research laboratory assay does not readily translate to human phenotype or drug response. Few assays are actually clinically validated for use in variant classification, and replication studies are rarely pursued. ClinGen's recommended approach to

assessing functional assays is a great start (8), though I suspect we will ultimately need even stricter approaches. I believe, however, that it will be the continued development of large databases of quality-controlled sequencing with matched deep phenotyping of populations and drug response data that will be required for next-level variant classification science.

7.3.2. Relationship between research genetic testing and clinical genetic testing. M.R.H.: Variants in genes are being identified all the time, though many do not have enough evidence for disease or disorder causality. It is unfortunate that laboratories are beginning to report variants as causative of a disease/disorder when causality is not apparent and is not supported by (sufficient) evidence. We need clinical research that accompanies basic research. What we are trying to do is understand the functionality of our basic biological systems, and we need basic researchers to have robust collaborations with the clinical community, with research funding mechanisms that encourage these collaborations.

E.L.: Submissions from ClinVar coming from both research and clinical labs is a double-edged sword. If I encounter a discrepancy in ClinVar, then I need to go in and look at it. In my experience, often that discrepancy is coming from a research lab. To be clear, I want to know if a research lab identified a variant and if they have evidence of pathogenicity. However, I would appreciate improved transparency around whether a submission is research or clinically based in origin.

7.4. Pressures on the Variant Classification Industry

7.4.1. Budget cuts and limitations on variant classification science. M.R.H.: Patient care is the most important casualty of limited budgets and resources, because reports are not going to be high quality in these situations. This directly compromises the quality of care the patient receives and the quality of the report itself.

C.M.S.: When health insurance payers do not reimburse or reimburse poorly, labs are forced to process more cases with fewer resources, increasing the probability for error. Automation cannot replace highly qualified variant interpretation staff who do a thorough evaluation of patient reports. Also, publication paywalls can impact classification by preventing access to the evidence, in which case either the evidence is not used in classification or the classification is based on the abstract alone, with no critical review of the evidence.

J.M.E.: The process of variant classification is not regulated in any meaningful way by CLIA or CAP, and "FDA-approved" NGS tests often have major portions of the test that are not actually regulated in a way that would safeguard against misclassified variants (23). Clinicians and health insurance payers have little to no insight into the variant classification quality of laboratories beyond what they hear in sales pitches. Decreasing the quality of variant classifications provides massive savings to laboratories with no consequence, so many labs cut corners. Patients, however, unknowingly pay with their lives due to misdiagnosis.

7.4.2. The pressure to upgrade variants of uncertain significance in the name of compassion. M.R.H.: If we consider a recessive germline case in which one pathogenic variant has been identified, the other variant is a VUS, and the phenotype matches the gene—indicating you are probably in the correct gene—then abstaining from delivering a therapeutic regimen is not right. Keep the classification of the variant as VUS—i.e., do not upgrade the variant to pathogenic just to deliver therapy—but do the compassionate action and deliver treatment. In cases such as this, we need a mechanism to track whether the patient responded to treatment and to give that information to the variant classification community. I suggest a clinical response category for ClinVar.

I.D.K.: If the data do not support a certain classification, then that is what we should communicate to the patient.

C.M.S.: I think compassion plays into variant classification in the sense that we are driven by the patients and a desire to help find them an answer. It's what keeps us up late at night reviewing publications in search of the cause of the patient's clinical symptoms.

7.4.3. The pressure to increase diagnostic yield. C.M.S.: The expectation of a higher diagnostic yield should actually add pressure to select the right patients for the test and not to overclassify variants as pathogenic for the sake of an appearance of a high diagnostic yield. High diagnostic yields are the result of a combination of selecting the correct patients for genetic testing and having a genetic test that can accurately detect the broadest spectrum of variant types (e.g., single-nucleotide variants, deletions and insertions, short tandem repeats, and copy number variants).

M.R.H.: It is nothing short of dangerous to tie health insurance reimbursement to positive test results, as has been piloted by some payers. Policies such as these are not a good idea and will negatively impact the practice of laboratory medicine. One consequence will be an increase in false positive rates.

8. CONCLUSIONS

The precision medicine era has been launched primarily due to the early successes of clinical genetics and genomics. On the analytical side of NGS testing, reducing false negatives is a major focus of improvement. The reduction in variant classification false positives should also be a major focus of improvement in the industry. Thankfully, some members of the clinical genetics and genomics community are working to improve variant classification methodologies. A new wave of powerful technologies and methods have been implemented by leading commercial companies and by public consortia. Incentives and funding to build the infrastructures needed to develop data-rich, evidence-based, quality-controlled, and orthogonally validated truth sets will provide the framework needed to truly move variant classification forward in both germline and somatic testing. As the field improves, clinical genetic variant classification science will become more accurate and will significantly expand its contribution to precision medicine.

DISCLOSURE STATEMENT

H.H.G., M.E.G., and J.M.E. are employed by the Center for Genomic Interpretation, which has the independent and nonprofit mission to drive quality in clinical genetics, genomics, and precision medicine. The center's salaries are paid by funds generated through advancing this mission. Therefore, the publication of this review supports the center's mission and will assist in garnering additional funding.

LITERATURE CITED

1. Abnizova I, te Boekhorst R, Orlov YL. 2017. Computational errors and biases in short read next generation sequencing. *J. Proteom. Bioinform.* 10:1
2. ACMG Lab. Pract. Comm. Work. Group. 2000. ACMG recommendations for standards for interpretation of sequence variations. *Genet. Med.* 2:302–3
3. Agarwala V, Khozin S, Singal G, O'Connell C, Kuk D, et al. 2018. Real-world evidence in support of precision medicine: clinico-genomic cancer data as a case study. *Health Aff.* 37:765–72

4. Genet Ambry. 2020. CustomNext-*Cancer*®. *Ambry Genetics*. **https://www.ambrygen.com/providers/genetic-testing/2/oncology/customnext-cancer**
5. Amendola LM, Muenzen K, Biesecker LG, Bowling KM, Cooper GM, et al. 2020. Variant classification concordance using the ACMG-AMP variant interpretation guidelines across nine genomic implementation research studies. *Am. J. Hum. Genet.* 107:932–41
6. Balmaña J, Digiovanni L, Gaddam P, Walsh MF, Joseph V, et al. 2016. Conflicting interpretation of genetic variants and cancer risk by commercial laboratories as assessed by the prospective registry of multiplex testing. *J. Clin. Oncol.* 34:4071–78
7. Bennett C. 2019. Precision medicine adopts cancer variant guidelines, but the work isn't over. *Clin. OMICs* 6:16–17
8. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, et al. 2019. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 12:3
9. Broad Inst. MIT Harv. 2020. *ATM: ATM serine/threonine kinase*. Genome Aggreg. Database (gnomAD), version 2.1.1. **https://gnomad.broadinstitute.org/gene/ENSG00000149311?dataset=gnomad_r2_1**
10. Broad Inst. MIT Harv. 2020. *Deletion (3 bases): 7-117199644-ATCT-A (GRCh37)*. Genome Aggreg. Database (gnomAD), version 2.1.1. **https://gnomad.broadinstitute.org/variant/7-117199644-ATCT-A?dataset=gnomad_r2_1**
11. Chen D. 2020. Myriad: using RNA testing in variant classification since 2015. *Myriad myRisk*, Jan. 2. **https://myriadmyrisk.com/using-rna-in-variant-classification**
12. Chen Z, Yuan Y, Chen X, Chen J, Lin S, et al. 2020. Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. *Sci. Rep.* 10:3501
13. ClinGen (Clin. Genome Resour.). 2020. FDA recognized human variant database. *ClinGen*. **https://www.clinicalgenome.org/about/fda-recognition**
14. ClinGen (Clin. Genome Resour.). 2020. Variant pathogenicity curation. *ClinGen*. **https://www.clinicalgenome.org/curation-activities/variant-pathogenicity**
15. Color. 2019. *Hereditary cancer test*. Fact Sheet, Color, Burlingame, CA. **https://static.getcolor.com/pdfs/providers/Color_Hereditary_Cancer_Test_Fact_Sheet.pdf**
16. Curtin C. 2020. Tumor, germline testing of cancer patients can give discordant results. *GenomeWeb*, Nov. 23. **https://www.genomeweb.com/cancer/tumor-germline-testing-cancer-patients-can-give-discordant-results**
17. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 36:e105
18. Dolinsky JS, Hruska KS, Pesaran T, Richardson ME, Klein RT, et al. 2017. Efforts toward consensus variant interpretation by commercial laboratories. *J. Clin. Oncol.* 35:1261–62
19. Duncavage EJ, Abel HJ, Pfeifer JD. 2017. In silico proficiency testing for clinical next-generation sequencing. *J. Mol. Diagn.* 19:35–42
20. Eggington JM, Bowles KR, Moyes K, Manley S, Esterling L, et al. 2014. A comprehensive laboratory-based program for classification of variants of uncertain significance in hereditary cancer genes. *Clin. Genet.* 86:229–37
21. Emami NC, Leong L, Wan E, Van Blarigan EL, Cooperberg MR, et al. 2017. Tissue sources for accurate measurement of germline DNA genotypes in prostate cancer patients treated with radical prostatectomy. *Prostate* 77:425–34
22. Esterling L, Wijayatunge R, Brown K, Morris B, Hughes E, et al. 2020. Impact of a cancer gene variant reclassification program over a 20-year period. *JCO Precis. Oncol.* 4:944–54
23. FDA (Food Drug Adm.). 2017. *CDRH'S approach to tumor profiling next generation sequencing tests*. Fact Sheet, FDA, Silver Spring, MD. **https://www.fda.gov/media/109050/download**
24. Fenton AT, Anderson EC, Scharnetzki E, Reed K, Edelman E, et al. 2020. Differences in cancer patients' and clinicians' preferences for disclosure of uncertain genomic tumor testing results. *Patient Educ. Couns.* 104:3–11
25. Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. 2014. Accuracy of next generation sequencing platforms. *J. Next Gener. Seq. Appl.* 1:1

26. Ganguly P. 2019. *NHGRI funds centers for advancing the reference sequence of the human genome*. News Release, Sept. 24, Natl. Hum. Genome Res. Inst., Bethesda, MD. **https://www.genome.gov/news/news-release/NIH-funds-centers-for-advancing-sequence-of-human-genome-reference**
27. Gao P, Zhang R, Li Z, Ding J, Xie J, Li J. 2019. Challenges of providing concordant interpretation of somatic variants in non-small cell lung cancer: a multicenter study. *J. Cancer* 10:1814–24
28. GeneDx. 2020. *OncoGeneDx: comprehensive common cancer panel*. Test Inf. Fact Sheet, GeneDX, Gaithersburg, MD. **https://www.genedx.com/wp-content/uploads/2013/09/B275.Comprehensive-Common-Cancer-Panel-TIS.v3-1_17_20.pdf**
29. GenomeDenmark. 2020. GenomeDenmark platforms. *GenomeDenmark*. **http://www.genomedenmark.dk**
30. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–22
31. Hagenkord J, Funke B, Qian E, Hegde M, Jacobs KB, et al. 2020. Design and reporting considerations for genetic screening tests. *J. Mol. Diagn.* 22:599–609
32. Harrison SM, Dolinksy JS, Chen W, Collins CD, Das S, et al. 2018. Scaling resolution of variant classification differences in ClinVar between 41 clinical laboratories through an outlier approach. *Hum. Mutat.* 39:1641–49
33. Hehir-Kwa JY, Marschall T, Kloosterman WP, Francioli LC, Baaijens JA, et al. 2016. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat. Commun.* 7:12989
34. Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Med*. 2:e124
35. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, et al. 2017. Guidelines for validation of next-generation sequencing–based oncology panels. *J. Mol. Diagn.* 19:341–65
36. Karam R, LaDuca H, Richardson ME, Pesaran T, Chao E. 2020. RNA-seq analysis is a useful tool in variant classification. *JCO Precis. Oncol.* 4:1226–27
37. Kehr B, Helgadottir A, Melsted P, Jonsson H, Helgason H, et al. 2017. Diversity in non-repetitive human sequences not found in the reference genome. *Nat. Genet.* 49:588–93
38. Kehr B, Melsted P, Halldórsson BV. 2016. PopIns: population-scale detection of novel sequence insertions. *Bioinformatics* 32:961–67
39. Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol*. 11:R116
40. Kerr ID, Cox HC, Moyes K, Evans B, Burdett BC, et al. 2017. Assessment of in silico protein sequence analysis in the clinical classification of variants in cancer risk genes. *J. Community Genet*. 8:87–95
41. Koboldt DC. 2020. Best practices for variant calling in clinical sequencing. *Genome Med*. 12:91
42. Landrum MJ, Kattman BL. 2018. ClinVar at five years: delivering on the promise. *Hum. Mutat.* 39:1623–30
43. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer genes. *Nature* 499:214–18
44. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, et al. 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*. 19:4–23
45. Lincoln SE, Yang S, Cline MS, Kobayashi Y, Zhang C, et al. 2017. Consistency of *BRCA1* and *BRCA2* variant classifications among clinical diagnostic laboratories. *JCO Precis. Oncol.* 1. **https://doi.org/10.1200/PO.16.00020**
46. Logsdon GA, Vollger MR, Eichler EE. 2020. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21:597–614
47. Ma X, Shao Y, Tian L, Flasch DA, Mulder HL, et al. 2019. Analysis of error profiles in deep next-generation sequencing data. *Genome Biol*. 20:50
48. Maretty L, Jensen JM, Petersen B, Sibbesen JA, Liu S, et al. 2017. Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature* 548:87–91
49. May A, Abeln S, Buijs MJ, Heringa J, Crielaard W, Brandt BW. 2015. NGS-eval: NGS Error analysis and novel sequence VAriant detection tooL. *Nucleic Acids Res.* 43:W301–5

50. Merker JD, Devereaux K, Iafrate AJ, Kamel-Reid S, Kim AS, et al. 2019. Proficiency testing of standardized samples shows very high inter-laboratory agreement for clinical next generation sequencing-based oncology assays. *Arch. Pathol. Lab. Med.* 143:463–71
51. Mersch J, Brown N, Pirzadeh-Miller S, Mundt E, Cox HC, et al. 2018. Prevalence of variant reclassification following hereditary cancer genetic testing. *JAMA* 320:1266–74
52. Mikhail FM, Biegel JA, Cooley LD, Dubuc AM, Hirsch B, et al. 2019. Technical laboratory standards for interpretation and reporting of acquired copy-number abnormalities and copy-neutral loss of heterozygosity in neoplastic disorders: a joint consensus recommendation from the American College of Medical Genetics and Genomics (ACMG) and the Cancer Genomics Consortium (CGC). *Genet. Med.* 21:1903–16
53. Moncur JT, Bartley AN, Bridge JA, Kamel-Reid S, Lazar AJ, et al. 2019. Performance comparison of different analytic methods in proficiency testing for mutations in the *BRAF*, *EGFR*, and *KRAS* genes: a study of the College of American Pathologists Molecular Oncology Committee. *Arch. Pathol. Lab. Med.* 143:1203–11
54. Morris B, Hughes E, Rosenthal E, Gutin A, Bowles KR. 2016. Classification of genetic variants in genes associated with Lynch syndrome using a clinical history weighting algorithm. *BMC Genet*. 17:99
55. MyriadGenet. 2020. *Myriad myRisk® Hereditary Cancer technical specifications*. Tech. Specif., Myriad Genet., Salt Lake City, UT. **https://s3.amazonaws.com/myriad-library/technical-specifications/myRisk+Hereditary+Cancer+Tech+Specs.pdf**
56. Natl. Cent. Biotechnol. Inf. 2019. *Invitae multi-cancer panel: performance characteristics*. Genet. Test. Regist., updated Oct. 15. **https://www.ncbi.nlm.nih.gov/gtr/tests/528909/performance-characteristics**
57. Natl. Cent. Biotechnol. Inf. 2021. *NM_000492.3(CFTR):c.1521_1523delCTT (p.Phe508delPhe)*. ClinVar, Access. VCV000007105.25, updated Jan. 16. **https://www.ncbi.nlm.nih.gov/clinvar/variation/7105**
58. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, et al. 2017. Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. *Genet. Med.* 19:1105–17
59. Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, et al. 2018. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* 8:10950
60. Pirooznia M, Kramer M, Parla J, Goes FS, Potash JB, et al. 2014. Validation and assessment of variant calling pipelines for next-generation sequencing. *Hum. Genom.* 8:14
61. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, et al. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* 29:1282–91
62. Pruss D, Morris B, Hughes E, Eggington JM, Esterling L, et al. 2014. Development and validation of a new algorithm for the reclassification of genetic variants identified in the *BRCA1* and *BRCA2* genes. *Breast Cancer Res. Treat.* 147:119–32
63. Reich D, Nalls MA, Kao WHL, Akylbekova EL, Tandon A, et al. 2009. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLOS Genet*. 5:e1000360
64. Richards CS, Aziz N, Bale S, Bick D, Das S, et al. 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.* 17:405–24
65. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, et al. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet. Med.* 10:294–300
66. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, et al. 2020. Technical standards for the interpretation and reporting of constitutional copy number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet. Med.* 22:245–57
67. Rosenthal ET, Bowles KR, Pruss D, van Kan A, Vail PJ, et al. 2015. Exceptions to the rule: case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin. Genet.* 88:533–41
68. Rusk N. 2018. The UK Biobank. *Nat. Methods* 15:1001
69. Shah N, Hou Y-CC, Yu H-C, Sainger R, Caskey CT, et al. 2018. Identification of misclassified ClinVar variants via disease population prevalence. *Am. J. Hum. Genet.* 102:609–19
70. Sherman RM, Forman J, Antonescu V, Puiu D, Daya M, et al. 2019. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* 51:30–35

71. Spence T, Sukhai MA, Kamel-Reid S, Stockley TL. 2019. The Somatic Curation and Interpretation Across Laboratories (SOCIAL) project—current state of solid-tumour variant interpretation for molecular pathology in Canada. *Curr. Oncol.* 26:353–60
72. Surrey LF, Oakley FD, Merker JD, Long TA, Vasalos P, et al. 2019. Next-generation sequencing (NGS) methods show superior or equivalent performance to non-NGS methods on BRAF, EGFR, and KRAS proficiency testing samples. *Arch. Pathol. Lab. Med.* 143:980–84
73. Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, et al. 2014. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.* 46:107–15
74. Vail PJ, Morris B, van Kan A, Burdett BC, Moyes K, et al. 2015. Comparison of locus-specific databases for *BRCA1* and *BRCA2* variants reveals disparity in variant classification within and among databases. *J. Community Genet.* 6:351–59
75. Yang S, Lincoln SE, Kobayashi Y, Nykamp K, Nussbaum RL, Topper S. 2017. Sources of discordance among germ-line variant classifications in ClinVar. *Genet. Med.* 19:1118–26
76. Yen JL, Garcia S, Montana A, Harris J, Chervitz S, et al. 2017. A variant by any name: quantifying annotation discordance across tools and clinical databases. *Genome Med.* 9:7
77. Zhao S, Agafonov O, Azab A, Stokowy T, Hovig E. 2020. Accuracy and efficiency of germline variant calling pipelines for human genome data. *Sci. Rep.* 10:20222