The Journal of Molecular Diagnostics, Vol. 17, No. 4, July 2015

2015 JMD CME Program

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

REVIEW

Annotation of Sequence Variants in Cancer Samples

Processes and Pitfalls for Routine Assays in the Clinical Laboratory

Lobin A. Lee, Kevin J. Arvai, and Dan Jones

From the Department of Pathology, Quest Diagnostics Nichols Institute, Chantilly, Virginia

CME Accreditation Statement: This activity ("JMD 2015 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint providership of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2015 CME Program in Molecular Diagnostics") for a maximum of 36 *AMA PRA Category 1 Credit(s)*™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication March 23, 2015.

Address correspondence to Lobin A. Lee, M.S., Department of Molecular Oncology, Quest Diagnostics Nichols Institute, 14225 Newbrook Dr, Chantilly, VA 20151. E-mail: lobin.a.lee@questdiagnostics.com.

As DNA sequencing of multigene panels becomes routine for cancer samples in the clinical laboratory, an efficient process for classifying variants has become more critical. Determining which germline variants are significant for cancer disposition and which somatic mutations are integral to cancer development or therapy response remains difficult, even for well-studied genes such as *BRCA1* and *TP53*. We compare and contrast the general principles and lines of evidence commonly used to distinguish the significance of cancer-associated germline and somatic genetic variants. The factors important in each step of the analysis pipeline are reviewed, as are some of the publicly available annotation tools. Given the range of indications and uses of cancer sequencing assays, including diagnosis, staging, prognostication, theranostics, and residual disease detection, the need for flexible methods for scoring of variants is discussed. The usefulness of protein prediction tools and multimodal risk-based or Bayesian approaches are highlighted. Using *TET2* variants encountered in hematologic neoplasms, several examples of this multifactorial approach to classifying sequence variants of unknown significance are presented. Although there are still significant gaps in the publicly available data for many cancer genes that limit the broad application of explicit algorithms for variant scoring, the elements of a more rigorous model are outlined. *(J Mol Diagn 2015, 17: 339—351; http://dx.doi.org/10.1016/j.jmoldx.2015.03.003)*

Germline Polymorphisms and the Risk of Cancer Development

Genetic differences in individuals include single nucleotide polymorphisms (SNPs), intragenic insertion and deletion polymorphisms (indels), and structural variants, such as copy number variations. These factors contribute to risk of cancer development and responses to therapy (pharmacogenomics). During tumor development, there is a complex interplay between somatic or acquired mutations in oncogenes, tumor suppressors, and epigenetic regulators and germline, or inherited, genetic variation.

Initial work on localizing genetic variants associated with cancer susceptibility focused on well-defined clinical syndromes, such as alterations of *TP53* in Li-Fraumeni syndrome,

Disclosures: All authors were employees of Quest Diagnostics at the time this review was composed. Quest Diagnostics offers sequencing assays commercially.

Current address of D.J., Ohio State University, Columbus, OH; of K.J.A., GeneDX, Gaithersburg, MD.

Copyright © 2015 American Society for Investigative Pathology and the Association for Molecular Pathology.
Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmoldx.2015.03.003

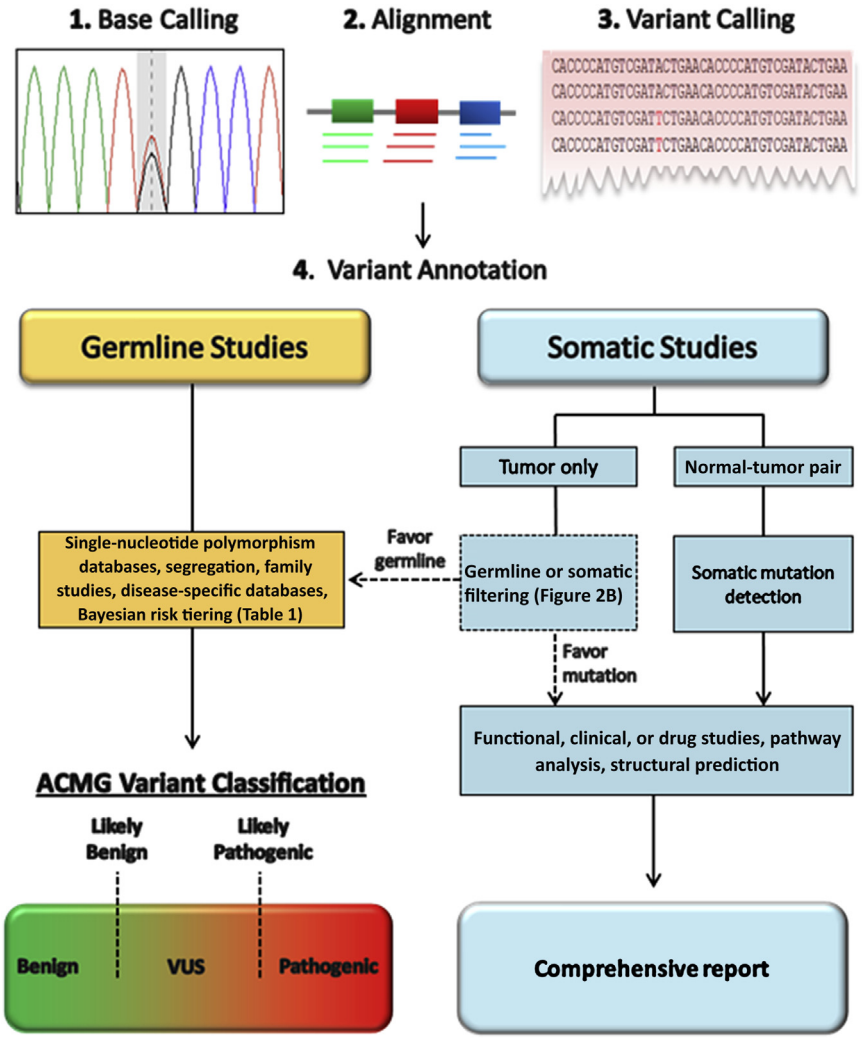

Figure 1 Variant analysis pipeline comparing germline and somatic annotation. Base-calling, alignment, and variant calling (steps 1 to 3) typically use a standard toolset, such as Samtools, Genome Analysis Toolkit (*https://www.broadinstitute.org/gatk*), Bowtie (*http://bowtie-bio.sourceforge.net*), and Burrows-Wheeler Aligner (*http://bio-bwa.sourceforge.net*). All websites were last accessed October 2, 2014. For variant annotation (step 4), the toolset and analysis parameters are less standardized. For germline studies, the American College of Medical Genetics and Genomics (ACMG) variant classification system provides guidance for interpretation. For somatic mutations in cancer samples, several tools are available for germline filtering if a normal and nonneoplastic reference sample is available for comparison, including MuTect (*http://www.broadinstitute.org/cancer/cga/mutect*), Varscan (*http://varscan.sourceforge.net*), and SomaticSniper.[7] When no reference sequence is available, rules for trimming germline calls must be applied before the somatic calls can be annotated for significance. VUS, variant of unknown significance.

BRCA1 and *BRCA2* in hereditary breast and ovarian cancer,[1] and the mismatch repair genes in Lynch syndrome.[2] The genes involved, typically tumor suppressors, were initially localized using linkage analysis and targeted DNA sequencing. The genetic changes observed in affected individuals include frameshift and nonsense mutations as well as inactivating mutations with loss of function linked to tumor initiation.

More recently, genome-wide association studies (GWASs), which compare variant profiles of diseased versus healthy individuals, have accelerated the rate of discovery of cancer-associated variants.[3] GWASs have identified many more cancer-associated variants in well-characterized cancer genes than family studies, including many missense mutations that have subtle or undetermined effects. They have also identified recurrent germline variants in genes whose association with carcinogenesis was not previously known. These include genes that serve core cellular functions, such as energy metabolism, chromatin maintenance, and protein translation.

Connecting a given variant to its phenotypic effect(s) is more difficult for GWASs compared with classic genetic analyses. Even when published data are available to support interpretation of a particular mutation call, GWASs have often had low reproducibility attributable to weak or incomplete phenotypic penetrance, bias against effects due to more common variants and covariants, and inadequate statistical power.[4] As a result, only approximately one-third of germline variant associations reach statistical significance across multiple studies.[5] Unsurprisingly, GWASs, massively parallel exome sequencing projects, and routine targeted next-generation sequencing (NGS) in clinical laboratories have resulted in many more variants of undetermined significance (VUS) in cancer-associated genes.[6]

The Central Role of the Analysis Pipeline

The first step in ensuring robust variant calling relies on accurate base-calling and alignment. To accomplish this goal, raw data from high-throughput sequencers are moved through an analysis pipeline to sequentially accomplish sequence alignment, read filtering, variant calling, and variant

annotation (Figure 1).[7] The initial steps may be accomplished by instrument vendor software, commercial or open source third-party software, or combinations. Guidelines from the College of American Pathologists and the New York State Department of Health (*http://www.wadsworth.org/labcert/TestApproval/forms/NextGenSeq_ONCO_Guidelines.pdf*, last accessed February 15, 2015) have benchmarked key data quality and analysis requirements for somatic mutation detection.[8]

Measuring sequence quality, read depth, and coverage are critical at each step of the pipeline to ensure adequate sensitivity and control for biases introduced by the nucleic acid quality, sequencing chemistry, assay design, and alignment software. Data quality cutoffs will differ by application, but greater read depths are needed for somatic mutation studies. This is especially true when suboptimal sample quality is expected because off-target reads and sequencing artifacts are more common. Other important quality measures include base and read quality filters, cutoffs for uniformity of target coverage, and maximum allowable strand bias for paired-end or bidirectional sequencing methods.

Data Sources and Classification Models for Germline Genetic Variants

Translation of germline variant calls into clinical decisions relies on proper annotation (Figure 1). There are now several public sources that catalog the frequency and population characteristics of germline variants. For SNPs, the International HapMap Project, the Exome Sequencing Project, and the 1000 Genomes Project report population-based data.[9–11] The online SNP catalogs Genevar (Sanger Institute)[12] and the Single-Nucleotide Polymorphism database (dbSNP; NIH database of germline variation[13]) house records for >100 million variants. The Database of Genomic Structural Variation (dbVAR) and the Database of Genomic Variants (DGV) catalog large-scale genomic variation (copy number variations), including large insertions, deletions, and inversions.[13,14]

Most DNA sequencing pipelines routinely query these sources, and the provided variant frequencies can be used to filter out commonly occurring changes. Presumed benign variants are typically regarded as those with minor allele frequencies (MAFs) >1% to 5%.[15] MAF segregation by race, provided by the Exome Sequencing Project, can be used if population demographics are relevant for tested patients. The assumption that major variants are of limited diagnostic utility may occasionally be erroneous but is essential for reporting large gene panels to narrow the number of variants requiring further analysis. However, most SNPs occur at MAFs under 0.5% (<1% of the population),[15] highlighting the difficulty of variant annotation.[16]

Most germline analysis pipelines also query the Human Genome Mutation Database, the Online Mendelian Inheritance in Man, the Clinical Genome Resource, and ClinVar to report clinical associations of the best-characterized pathogenic germline variants.[13,17] At this time, curated gene-specific databases cover only a few cancer-associated genes (eg, *TP53* and *BRCA1/2*). Therefore, locally curated variant databases are essential for reporting and identifying significant co-occurrences with other variants.

Having excluded common (presumed nonpathogenic) variants and highlighted pathogenic ones, indeterminate calls must be scored (Figure 1). The American College of Medical Genetics and Genomics (ACMG)[18] and the International Agency for Research on Cancer (IARC)[19] have released guidelines on germline sequence variant interpretation to promote standardized nomenclature. The ACMG and IARC systems rely on lines of evidence (LOE) to stratify germline variants into tiers from nonpathogenic (benign) to definitively pathogenic. LOEs include linkage and segregation data from pedigree analysis and family studies, population-based data on relative risk, clinical correlations, *in vitro* functional studies, predictions of protein structural effects, evolutionary conservation, and frequency distributions (Table 1). Most laboratories do not categorically score variants for each factor but rely most heavily, for cancer genes, on relative risk and clinical correlations.

Approaches to the VUS Problem for Germline Variants in the NGS Era

The best-studied cancer susceptibility genes, particularly *BRCA1, BRCA2,* and the Lynch syndrome–associated mismatch repair genes, have publicly available and curated databases.[20,21] These efforts have helped to limit reporting in routine clinical assays to those variants for which there is strong suspicion for an inherited basis for a patient's tumor. Reporting a poorly characterized VUS, even in a well-studied cancer gene,[22] can have detrimental consequences. The lack of clear guidance may lead some patients to avoid beneficial standard therapies, whereas others may opt for unnecessary procedures. These unintended effects can be exacerbated if the testing has reproductive or screening implications for family members.

In multigene cancer susceptibility panels, tens to hundreds of genes that are not as well annotated as *BRCA1* are now being reported. In these panels, VUS calls have increased exponentially, encompassing variants for which there is not yet sufficiently strong evidence of clinical and/or functional significance, those with limited population frequency data, or in which the existing data are contradictory. Given the limited association of some of these genes with genetically defined cancer syndromes, annotation of large NGS panels will require different approaches.

Interpretation of a VUS call can be based, in part, on the pretest probability of a positive test result, as determined by demographic and/or clinical risk factors. Effective implementation of this type of Bayesian or tiered-risk approach depends on the availability of reliable, correlative data. Given

Table 1 Germline and Somatic Variants Compared and Contrasted

Germline sequence variants	Somatic mutations
General features	
Disease association: Single gene and single disease (eg, *CFTR* in cystic fibrosis), single gene and multiple diseases, multiple genes and single disease (Lynch syndrome)	Mutation specificity: Single gene and single cancer type (eg, *BCR-ABL1* in CML), single gene and multiple cancer types (eg, *KRAS*), multiple genes and single cancer (*CALR, JAK2* V617F in MPN)
Variant effects: Directly pathogenic (dominant or recessive), interacting effects, linkage to altered gene or locus	Mutation effects: Tumor initiation, promotion, outgrowth, metastasis, progression, or therapy resistance
Allele frequency: Linked to patient population and/or racial group	Mutation frequency: Highly variable based on oncogenicity, tumor type, prior treatment(s)
Level: Usually present in all cells and detected at ~50% or ~100% levels	Level: Variable due to percentage of tumor present, gene copy number (ploidy), and subclonal occurrence
Retained throughout disease course except if locus is deleted	Mutation persists, lost, and/or reacquired due to tumor evolution or treatment selection
Clinical uses: Disease predisposition, family risk and reproductive guidance	Clinical uses: Diagnosis, prognosis, recurrence monitoring therapy (Table 2)
Challenges for annotation and reporting	
Limited clinical correlations for poorly penetrant or uncommon variants	Distinguishing somatic from germline variants if no normal or nonneoplastic reference sequence
Linking variant effect to phenotype	Linking mutation to tumor category or outcome
Inferring magnitude of variant effect(s)	Linking mutation to treatment response
Strong evidence for pathogenicity or oncogenicity	
Disease and variant segregation studies	Clinical studies linking mutation to therapy outcome or prognosis in multivariate models are few
Case-controlled studies comparing prevalence of variant in affected versus healthy populations	Highly powered correlative studies on mutation frequency by tumor type and/or clinical stage
Functional studies reveal damaging effect of variants	*In vitro* animal or cell studies revealing transforming or tumor suppressive effects of a mutation
Weaker or inferential evidence for pathogenicity or oncogenicity	
No occurrence or rarity in SNP databases	Frequency in somatic mutation databases and literature
Family studies to trace inheritance patterns and identify *de novo* changes	Mutation burden tracking with disease response and recurrence
Bayesian risk assessment based on pre-test probability in an individual patient	Bayesian risk assessment based on other high-risk clinicopathologic features (age, stage, histology)
Co-occurrence with known pathogenic variants reduce risk	Co-occurrence with more definitive disease-defining, prognostic or response-predicting mutations
cis- or trans-inheritance patterns (eg, VUS on the same allele as a known pathogenic would favor benign classification of VUS)	Pathway analysis to detect complementing mutations or those that are known to be mutually exclusive with variant detected
Computational tools predict pathogenic variants	Computational tools predict contribution of variant to disease

CML, chronic myelogenous leukemia; MPN, myeloproliferative neoplasm; SNP, single nucleotide polymorphism; VUS, variant of unknown significance.

the need for well-annotated reference patient populations, it is most easily used in highly structured clinical programs and can be difficult to implement in a reference laboratory setting. Below, we describe how such multivariate approaches can be applied to the even more complex task of annotating somatic sequence variants in cancer samples.

The Types of Somatic Sequence Variants in Cancer Samples

Genetic changes that arise during the development of a tumor are termed somatic mutations and possess commonalities and differences with germline changes (Table 1). Acquired somatic mutations in cancer cells are propagated through clonal expansion from founder cancer stem cells or tumor subpopulations. If a given genetic change promotes tumor development, it is regarded as a driver mutation and is typically retained during the disease course.[23] Such pathogenic mutations are typically classified as gain-of-function changes in tumor-promoting oncogenes or loss-of-function changes in tumor suppressor genes. The latter effects can also be produced by deletion of the entire gene. Other cancer-associated mutations have dominant-negative or hypofunctional effects that do not fit this oncogene and tumor suppressor duality. The latter mutations are common in genes that serve core metabolic functions or those that regulate epigenetic properties, such as histone and DNA methylation or acetylation and posttranscriptional RNA or protein modifications.

Once a tumor becomes established, additional mutations that were present in the selected abnormal cell population but that are not integral to tumorigenesis can arise as passengers. The differentiation of driver from passenger is not

Table 2 Uses of Multigene Sequencing Assays in Cancer Testing

Examples	Analytic considerations
Diagnostic classification and subclassification	
Molecular subclassification of hematologic malignant tumors (eg, *NPM1*- and *CEBPA*-mutated normal karyotype AML)	Strength of association between a specific mutation and tumor subtype is highly context dependent
Cancers presenting with unknown primary	Most tumor types lack mutations with high diagnostic specificity
Identifying high-risk molecular changes in tumor with typical histologic features	Separate sampling of histologically divergent areas may be necessary for molecular analysis
Clonality assessment	
T-cell and B-cell clonality by TCR and BCR repertoire profiling	May change definition of clonality in many lymphoid malignant tumors due to increased sensitivity
Differentiating reactive hyperplasia from early-stage neoplastic lesions	The full range of normal findings in hyperplastic lesions must be fully understood
Detecting recurrent tumor in small or limited biopsy specimens in which histologic analysis is compromised	May not reflect clinically significant findings if molecular results are interpreted in the absence of definitive microscopic findings
Prognostication	
Multimutation models of outcome within a standard clinicopathologic risk group	Association between a mutation and outcome linked to specific data sets, may not be generally applicable
Replacement for multimodal testing for upfront risk assessment for decision to treat (eg, CLL prognostic models)	Comparability of mutation and gold standard nonmutation testing models (eg, FISH panels in CLL) needs to be established
Theranostics	
Multigene hotspot panels assessing actionable mutations for treatment options in refractory or relapsed disease	Many mutations identified will be lightly annotated and not linked to well-established therapies, limited clinical trials options
Multimutation models to select patients for neoadjuvant or maintenance therapies	Interpretation dependent on strength and relevance of model data
Deep sequencing to detect emerging resistance mutation (eg, *ABL1* kinase domain mutations in CML)	Limited guidance on patient management in such preemergent low-level mutation settings
Minimal residual disease assessment	
Molecular fingerprinting tumors at diagnosis for followup monitoring	Stability of mutation profile during disease course will depend on oncogenic strength and treatment effects

AML, acute myeloid leukemia; BCR, B-cell receptor (*IGH, IGK*); CLL, chronic lymphocytic leukemia; CML, chronic myelogenous leukemia; FISH, fluorescence *in situ* hybridization; TCR, T-cell receptor genes (ie, *TCRG, TCRB*).

always clear-cut because many somatic mutations have cooperating and subtle effects on tumor growth. This is particularly true for epigenetic regulators, such as *TET2*.

Tumor evolution, often represented as a linear progression in historical models, has now been found to have complex branched patterns in cancers.[24] Complex tumor progression patterns result in different secondary mutations present in various tumor subclones and highlight the importance of adequate tumor sampling and appropriately sensitive mutation detection techniques. The level of a particular mutation has important implications for interpretation (Table 1). Tumor-associated aneuploidy and genomic instability often produce unpredictable increases or decreases in the copy number of a mutated gene. Finally, some driver mutations in early-stage tumors can be lost as the neoplasm evolves and spreads, whereas others may be required for tumor persistence (also known as oncogene addiction).

Assay-Specific Goals in Annotation of Somatic Variants in Cancer

Similar to germline variant annotation, approaches to somatic variant annotation in cancers differs based on type of assay (full exome versus hotspot or targeted panels) and assay goal(s). Common indications for mutation detection include establishing clonality (ie, differentiating a reactive from neoplastic process), tumor subclassification, prognostic risk stratification, therapy response prediction (theranostics), and minimal residual disease assessment (Table 2).

NGS clonality assays include T-cell receptor (TCR) and B-cell antigen receptor (BCR) profiling in lymphoproliferative disorders,[25,26] identifying hematologic neoplasms in patients with blood cell abnormalities[27] and distinguishing atypical hyperplasia from early-stage precancer lesions.[28] Clonality analysis by mutation profiling also shows promise in helping diagnose tumors in limited specimens. In these types of assays, any somatic mutation can potentially serve as a useful marker to distinguish clonal from reactive cell expansion. Two tasks of annotations for these assays are distinguishing normal baseline levels of mutation from disease-associated changes and distinguishing germline from somatic variants if no reference (nontumor) material is available for comparison.

The use of mutation patterns to subclassify neoplasms is most common currently for hematologic tumors. For acute myeloid leukemia (AML), the latest classifications have incorporated several gene mutations along with chromosome changes into the diagnostic paradigm.[29–32] In solid

tumors, mutation profiling integrated with microarray genomics can identify molecular subgroups that cross histologically defined categories. An important example is the identification of *TP53*-mutated subgroups in both poorly differentiated and more typical histologic variants of breast, colon, and endometrial cancer.[33] An important caveat is the propensity of large gene panels to produce unexpected combinations of mutations that may lead to misleading interpretations because of the inevitable false-discovery rates that occur. Multigene mutation profiles also have some promise in identifying the originating tissue source for cancers presenting with unknown primary.[34]

If a sequencing assay is used for risk stratification, annotation must be tied to an underlying data set with strong statistical power and a similar clinical management strategy as the target population. This type of assay has gained the most traction in hematologic tumors. For example, the prognostic impact of different mutations in cytogenetically normal AML led to the adoption of *NPM1*-mutated and *CEBPA*-mutated risk categories in the 2008 World Health Organization classification.[35] However, treating all mutations within a gene as equal with regard to prognostic impact is insufficiently sophisticated. Especially for inactivating mutations, mutation pattern is usually relevant, such as the finding that only dual but not single mutations in *CEBPA* encode favorable prognosis in AML.[36]

For theranostic indications, annotation focuses on identifying important driver or resistance mutations that can guide therapy decisions. Examples include defining the inhibitor response pattern of specific *ABL1* kinase domains mutations associated with imatinib-resistant chronic myelogenous leukemia (CML),[37] categorizing which *EGFR* mutations in lung cancer can predict response to EGFR-directed kinase inhibitors[38] and detecting activating *BRAF* V600E/K mutations that qualify patients for use of BRAF inhibitors in melanoma.[39] However, relatively few mutation-specific targeted therapies have been found to correlate with outcome in clinical studies, and even fewer agents have been approved for clinical use based on mutation result.

Nonetheless, NGS theranostic panels comprising tens to hundreds of genes are now routinely used, particularly for relapsed or refractory solid tumors for which off-label or compassionate use of targeted agents is more common. Sequencing reports often include several potentially actionable variant calls, the types of therapeutic agents that might be used, and the currently open clinical trials for which a patient may qualify. The mutation-target associations reported are often based on *in vitro* correlative data rather than clinical trials.

A promising application of mutation profiling is the detection of minimal residual disease or low levels of potentially resistant subclones following targeted therapy. NGS assays can have the widest quantitative range of any routine laboratory platform, with reproducible sensitivities down to the single molecule level. These sensitivities are comparable or superior to flow cytometry and clone-specific real-time PCR.[40,41] NGS assays looking for mutation levels of $\leq$1% are regarded as deep sequencing and require upward of 1000X coverage.[42] Deep sequencing may even provide important information for solid tumors at diagnosis.[43] Mutation read percentage(s) can provide an accurate estimate of the size of the residual tumor present, although tumor aneuploidy can occasionally lead to imprecise quantitation. If mutation assays are to be used for minimal residual disease applications, however, the stability of specific mutations during the disease course needs to be established.[44]

Tools and Approaches for Annotating Somatic Variants in Cancer

For each application above, the read depth required, extent of sequencing performed, and need for confirmatory methods vary. These considerations include the expected ratio of normal and neoplastic cell populations, any requirement to identify subclonal mutations, the precision of quantitation needed, and expected off-target or poor quality reads if heavily fixed or degraded samples are to be tested. For diagnostic and theranostic applications seeking 5% allele sensitivity, most laboratories strive for read depths of 200 to 500 per region. A recent analysis of tumor and cell line genomic data sets revealed that detecting mutations at 5% allele frequency with 99% sensitivity requires sequencing depth of at least 200×.[45]

A variety of data analysis tools and approaches are used for accurate and clinically relevant variant annotation. An initial step for all studies with paired normal and nonneoplastic sequences is to distinguish germline from cancer-associated changes. Several well-constructed software tools are available for this task (Figure 1). There are comparatively fewer tools available to help define functional and clinical significance of somatic variants or to distinguish somatic from germline calls when no normal reference (nontumor) sequence is available. This latter task arises with hematologic indications and small biopsy samples in which normal and neoplastic tissue cannot be dissected for separate analysis.

A major challenge for sequence alignment tools in somatic NGS assays has been the frequent presence of highly variably sized cancer-associated intragenic insertions and deletions. Sequence reads with large indels can be inadvertently removed during pipeline analysis due to unresolved alignment or overly rigorous quality filters. They may also align incorrectly, leading to apparent false variant calls surrounding the indels. Applying additional realignment tools before quality filtering (eg, the IndelRealigner in Genome Analysis Toolkit or Pindel) is the most common solution. Incorporation of multiple tools boosts the ability to reliably call intragenic indels in cancer samples.[46]

Given the limited case numbers in most cancer studies, the breadth and accuracy of somatic mutation databases have lagged behind those available for germline changes. The Catalog of Somatic Mutations in Cancer (COSMIC version 71, *http://cancer.sanger.ac.uk/cancergenome/projects/cosmic*, last accessed November 4, 2014) is the most important

manually curated repository of published cancer mutation studies, providing systematic nomenclature for calls based on the cDNA sequence change, mutation frequencies by tumor site, literature linkouts, browser visualization tools, and access for pipeline queries.[47] A major deficiency of COSMIC has been the misassignment of some SNPs as mutations, principally due to the absence of normal reference sequences in many of its linked studies. Although COSMIC includes information on whether cited studies included a normal reference, uninformative and conflicting publications can still lead to misclassification. For the *TET2* gene discussed below, COSMIC contains at least 10 entries for mutations that are reported as normal variants in dbSNP and have been confirmed as germline changes in published studies.

The Cancer Genome Atlas and the International Cancer Genome Consortium are public databases that provide clinically linked annotation of somatic mutations. Although these databases have fewer tumor samples compared with the aggregated data in COSMIC, they provide a better model for annotation given their standardized data collection methods. Germline variant databases may also be used to annotate somatic calls because some important cancer mutations occur as somatic and germline changes.[48]

Distinguishing Somatic Mutations from Germline Variants in the Absence of a Reference Sequence

Cancer sequencing assays have been designed to establish diagnoses in early-stage lesions or limited samples, track disease levels after treatment, and subclassify tumors. All rely on accurately filtering out germline variants. Given the abundance of low frequency and previously undescribed SNPs,[10] this process can be difficult when no nonneoplastic reference sequence is available.

Mutations in *TET2* are now commonly used to help in the diagnosis and classification of myelodysplastic syndrome (MDS) and myeloproliferative neoplasm (MPN).[49] In this indication, somatic mutation profiling is an adjunct test when there is clinical suspicion and abnormal blood cell counts but a definitive diagnosis cannot be established based on other laboratory testing, such as flow cytometry and cytogenetic analyses.[35,50,51] In this follow-up or postbiopsy setting, there is typically only a blood or bone marrow sample available for sequencing. Therefore, careful annotation of the sequencing data is required to distinguish SNPs from somatic missense mutations. We illustrate below the various techniques that can be used to accomplish this task for *TET2*.

TET2 as an Example of a Highly Polymorphic Cancer—Associated Gene

TET2 belongs to the TET family of epigenetic regulatory enzymes that convert 5-methyl-cytosine to 5-hydroxymethyl-cytosine and coordinately regulates expression in many genes through global and site-specific DNA methylation changes.[52] Somatic or acquired *TET2* mutations occur at high frequency across a spectrum of myeloid and lymphoid malignant tumors[53] and are transforming in animal studies and cell line models[52,53] but also occur as an age-related phenomenon in hematopoiesis without overt leukemic changes.[54]

Most pathogenic mutations in *TET2* result in complete or partial loss of function, evidenced by attenuation of 5-hydroxymethyl-cytosine in leukocyte DNA from patients with *TET2*-mutated blood cells compared with those in unaffected populations.[53] These mutations are commonly frameshift and nonsense or terminating mutations occurring anywhere within its 2002 amino acids but also include many different missense mutations that are presumed to have hypofunctional effects. However, fully cataloging pathogenic missense *TET2* mutations has proven difficult because an ever-increasing number of SNPs are reported (Figure 2A) and normal/tumor-paired testing has been performed for only a small number of *TET2* variants in published series.[55,56] Although somatic mutation and SNP databases may be incomplete and even occasionally inaccurate, filtering calls with MAF >2% to 5% helps to narrow the number of variants needing annotation.

Inferring Somatic Status from Allele Frequency and Mutation Level

Because most NGS methods produce highly accurate estimates of allele frequency, read percentage is also helpful in distinguishing somatic from germline calls, especially if there is an estimate of percentage tumor involvement in the sample by other methods. Germline SNPs should be detected at or close to heterozygous (50%) or homozygous (100%) levels. In contrast, somatic mutations can occur at any level, depending on the amounts of neoplastic and normal cells present in the sample. However, use of this finding to infer somatic status needs to be interpreted with caution because ploidy changes in the tumor can influence apparent SNP allele frequencies. Correlation of mutation results in parallel genomic microarray or conventional cytogenetic studies can also assist in this distinction.

Distinguishing Variants by Domain Mapping and Protein Prediction

Mutagenesis studies and radiographic crystallography of the TET2-DNA complex have identified several conserved functional domains (CFDs) that are critical to TET2 function, including zinc-chelating sites and catalytic domains.[57] Missense mutations in *TET2* tend to cluster in these CFDs (Figure 2A), which can help classify variants based on their location.[55,56] Additional support for the pathogenicity of *TET2* missense changes can be obtained from cell lines transfected with mutagenized *TET2,* which reveals significantly decreased enzyme function when mutated at key residues, such as H1302Y and D1304A.[53] However,

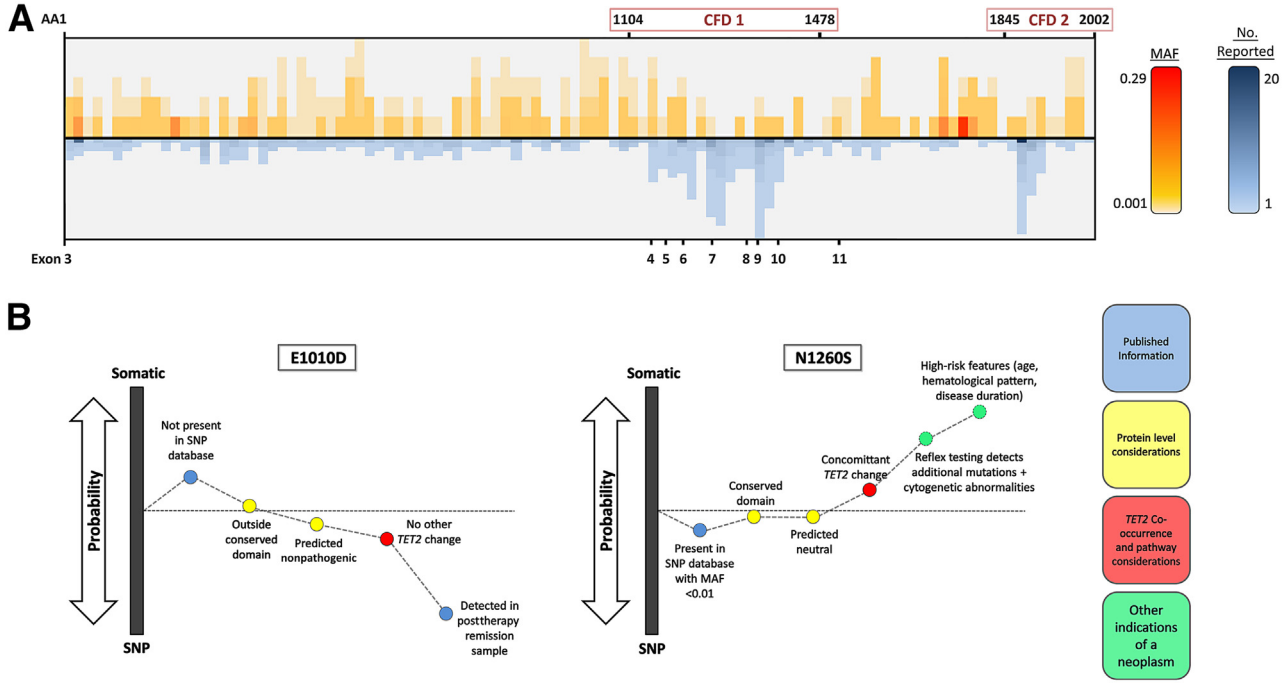

Figure 2 Variant classification of the *TET2* gene. **A:** The distribution of reported single-nucleotide polymorphisms (SNPs) and somatic missense mutations in the *TET2* gene. The coding region of the *TET2* gene is shown with exons; amino acid number and conserved functional domain (CFD) 1 and CFD 2 are indicated. SNPs reported for the gene in the Exome Sequencing Project are graphed above the center line in bins of 19 amino acids. The height above the center line reflects the cumulative frequency of variants in that region; minor allele frequency (MAF) for the most common SNPs is colored according to the scale on the right. The bottom half depicts the distribution of missense mutations reported in COSMIC version 71 (last accessed November 4, 2014) with distance from the center line representing the total number of changes in each bin and color representing the frequency of specific mutation calls. Any mutation also present in the above SNP databases is omitted. **B:** Distinguishing SNPs from somatic mutations in *TET2* using integrated lines of evidence. The *TET2* E1010D variant is not present in SNP databases so a somatic mutation may be indicated. However, localization outside the CFDs, nonpathogenic consensus by computational tools, and no additional abnormalities in *TET2* will make a SNP more likely. For the *TET2* variant N1260S, presence at low frequency in SNP databases but location inside a conserved domain makes classification difficult. However, if testing identifies other *TET2* mutations and/or complementing mutations in other genes, the probability of N1260S representing a somatic mutation will increase.

germline SNP and VUS calls also occur within these two CFDs, including more than 30 nonsynonymous entries in dbSNP (build 142), and somatic changes can occur outside CFDs.[58,59]

Domain mapping can be supplemented by computational tools that model the effects of missense variants *in silico*.[60] In general, these tools classify protein sequence variants on a continuum from benign to damaging and include phylogenetic approaches leveraging multiple sequence alignments of evolutionary homologs (eg, SIFT, *http://sift.jcvi.org*, last accessed October 10, 2014), phylogenetic-independent strategies that analyze biochemical and physiochemical patterns, and combinations of these approaches (eg, Polyphen-2, *http://genetics.bwh.harvard.edu/pph2*, last accessed October 10, 2014). Multiple sequence alignment methods assume a relationship between pathogenicity and evolutionary conserved residues and have found 72% to 82% accuracy for some cancer genes.[61] For *TET2*, SIFT and Polyphen-2 have a reduced accuracy outside conserved domains, a task for which a Fourier transform physiochemical algorithm has proven more effective.[62] Applying these tools in tumor samples assumes that pathogenic mutations are synonymous with cancer-associated changes, which may be inaccurate. Broadly applicable approaches specific for predicting oncogenicity for somatic mutations have not yet been developed. Therefore, in the absence of these tools, using multiple prediction algorithms can improve predictive power.[63]

Co-Occurrence and Pathway Analysis

For germline variants, the presence of a definitive pathogenic mutation tends to reduce the pathogenicity score of a co-occurring VUS. The opposite is often the case for somatic mutations in cancer where the co-occurrence of an uncharacterized missense variant along with a definitive oncogenic mutation in the same gene can boost significance. This is particularly true for genes in which biallelic mutations of different strengths (eg, one inactivating, one hypofunctional) are a predominant mechanism of transformation, such as with *TP53*, *TET2,* and *CEBPA*. For *TET2*, cancer samples often have one inactivating mutation and additional missense substitutions that would be predicted by the tools above to have subtle or hypofunctional effects.

A related tool is pathway analysis in which genes that are complementary are frequently mutated together, whereas

those that act along the same pathway or in the same complex are not. In hematopoietic malignant tumors, *TET2* mutations typically co-occur with mutations in the epigenetic regulators *DNMT3A* and *EZH2* but are mutually exclusive with *IDH1* and *IDH2* mutations.[64] The frequency of complementing mutations also increases as tumors progress.

For samples with especially critical indications, reflex testing with a larger sequencing panel or genomic and SNP microarray also represents a viable strategy for resolving indeterminate calls. As knowledge of mutational patterns characteristic for specific cancer types increases, the pattern of co-occurring alterations in other genes can also help resolve the nature of indeterminate calls.[65] For *TET2* profiling in myeloid neoplasms, these complementary molecular studies might include chromosome or genomic microarray analysis and sequencing for mutations in other genes, such as *ASXL1, EZH2,* and *RUNX1,* all of which co-occur with *TET2* mutations in MDS and MPN.[66]

Use of the Full Bioinformatics Toolbox to Help Resolve Indeterminate Calls

In the absence of any widely acceptable model for scoring somatic variants, individual laboratories currently design their own customized approaches to weighing lines of evidence. As an example, we present an approach to this problem using several VUSs in *TET2*. At this time, the significant gaps in the variant frequencies, structural data, and clinical literature for *TET2* (or any other polymorphic cancer gene) preclude an automated or explicit scoring algorithm. However, integration of the bioinformatics tools and multiple LOEs described above can be used effectively on a call-by-call basis to determine the likelihood of somatic versus germline origin.

For example, consider a blood sample submitted for MDS workup in which the E1010D missense change is detected in *TET2* at 46% frequency (Figure 2B). This variant has been uncommonly reported in MDS studies without normal reference samples but not in the SNP databases, favoring somatic origin. However, given the close to heterozygous level and the location of this amino acid outside both *TET2* CFDs, germline origin is possible. With the use of the prediction tools SIFT, Polyphen-2, Align-GVGD (*http://agvgd.iarc.fr*, last accessed October 10, 2014) and Mutation Assessor (*http://mutationassessor.org*, last accessed October 10, 2014), a nonpathogenic consensus score is returned. Germline status would be confirmed by its detection in a posttherapy remission sample, as has been seen in prior studies.[67]

In contrast, consider another blood sample submitted for MDS workup where *TET2* N1260S is detected in 45% of the reads (Figure 2B). This call has been rarely reported in the SNP databases (MAF <0.01) and is present at close to heterozygous levels. Because the amino acid lies inside a CFD, either a somatic mutation or an SNP would be considered. Using the functional predictors above, a neutral consensus score is returned. However, the identification of a *TET2* truncation in the same sample raises the probability that N1260S is an acquired change that produces a hypofunctional protein.

These results for any given call might then be combined with other mutation data, sample (eg, tumor type and grade), and patient features (eg, age and stage) to generate a Bayesian-type risk score for the variant, as has been performed for germline calls in *BRCA1* and *BRCA2*.[68] For *TET2* calls in AML, MDS, and MPN samples, if pathogenic mutations in other leukemia-associated genes or characteristic cytogenetic changes are identified, this approach might affect the probability of an equivocal variant call being somatic (Figure 2B).

Interpreting the Significance of Somatic Mutations

Once germline calls have been excluded, the task of reporting the significance of definitive somatic variants is highly context dependent. Unlike the ACMG five-tier approach to germline variant calls, no such simple classification is possible or even warranted for somatic mutation in cancers. For theranostic assays in which the goal is treatment guidance, mutations are often scored in three tiers as actionable, potentially actionable, or of unknown therapeutic significance. The confidence level for the score given is based on the strength of the underlying evidence obtained through the categories below.

Treatment Response in Clinical Studies

A highly significant criterion for an actionable theranostic mutation call is one or more well-powered studies linking that specific mutation or a highly similar change to outcome after targeted therapy.[69] Such studies have formed the basis of drug approvals linked to the presence or absence of a mutation and/or have been incorporated into treatment guidelines.[70] However, effects of mutation level on outcome and the variations in analytical sensitivity of the mutation detection methods[71] may be confounding variables. Given the relative rarity of any specific amino acid mutation change in even large studies, meta-analyses combining multiple studies can be used to refine the interpretation.[72] Overrepresentation or higher than expected prevalence of a specific mutation in a particular response group in a study designed for another purpose may be used as weaker evidence of a potentially actionable call.

Animal and Cell Line Models

Animal and *in vitro* studies can help establish the pathogenic nature of specific mutations and their treatment response profile to specific drugs. Transgenic and knockout

animal models in which the target cancer develops in the appropriate timeframe after introduction are the strongest lines of evidence for oncogenicity. Less definitive but still useful are cell line sensitivity studies in which the effects of different drugs on a target gene engineered with a range of clinically relevant mutations are tested *in vitro*. This method has been used to map the response of specific kinase inhibitors on various *ABL* kinase domain mutations.[73] This *in vitro* sensitivity data have then been successfully used to select effective alternate kinase inhibitors (principally nilotinib and dasatinib) for treatment of imatinib-resistant CML.[37]

Protein Prediction Tools

For some targetable genes, there is more information on expected behavior of observed mutations on drug response in the biochemical and structural biology literature. Computational tools that assess the functional effect of amino acid changes in proteins can give predictive information on a particular mutation (as described in *Distinguishing Somatic Mutations from Germline Variants in the Absence of a Reference Sequence*). Protein databases, such as UniProt,[13] provide highly structured and searchable tools that link amino acid regions to functional and drug-binding sites. If a particular mutation maps to the areas of the protein predicted to mediate drug response, a potentially actionable score may be achieved.

Conclusion

The elements of standardized practices for accurate analysis and annotation of somatic mutations in cancer have been presented here and elsewhere.[74] but formal guidelines have not yet been promulgated. Relying on public databases is not sufficient at this time for annotation of most oncology assays. This is partly because of the constant occurrence of novel variants, the incomplete nature and inevitable classification errors in the SNP databases and COSMIC,[47] and the indeterminate somatic or germline status of some variants (eg, *EGFR* T790M).[48] A recent meta-analysis of 49 cancer genomics studies found that some were deficient in their analysis pipeline and that most did not adequately explain their processes for filtering germline calls and assigning significance to variants.[75] These deficiencies will limit the utility of these promising new technologies and may lead to poor clinical decision-making. More fundamentally, the assumptions and goals of annotation for different types of cancer sequencing assays have not yet been rigorously explored.

The molecular diagnostics community will need to continue to advocate for adoption of annotation guidelines to realize the full diagnostic potential and clinical utility of multigene NGS assays. Most urgently, the parameters for an actionable call in theranostic sequencing assays need formulation. It is likely that no single scoring system or rules-based approach will be sufficient for determining significance of somatic variants calls in all cancer applications, but it is imperative that more consistent practices be developed. As the sophistication of sequencing technologies has quickly increased, so too must the manner in which clinicians and investigators use these tools.

Acknowledgments

The authors thank Sean Caruthers and Ivy Tran for assistance with *TET2* data analysis.

References

1. King MC, Marks JH, Mandell JB: Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science 2003, 302: 643−646
2. Lynch HT, de la Chapelle A: Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet 1999, 36:801−818
3. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007, 447:661−678
4. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. Nature 2009, 461:747−753
5. Dong LM, Potter JD, White E, Ulrich CM, Cardon LR, Peters U: Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. JAMA 2008, 299:2423−2436
6. McClellan J, King MC: Genetic heterogeneity in human disease. Cell 2010, 141:210−217
7. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L: SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 2012, 28:311−317
8. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Grody WW, Hegde MR, Hoeltge GA, Leonard DGB, Merker JD, Nagarajan R, Palicki LA, Robetorye RS, Schrijver I, Weck KE, Voelkerding KV: College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. Arch Pathol Lab Med 2015, 139:481−493
9. The International HapMap Project. Nature 2003, 426:789−796
10. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM: Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012, 337:64−69
11. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: An integrated map of genetic variation from 1,092 human genomes. Nature 2012, 491:56−65
12. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET: Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 2010, 26: 2474−2476
13. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2014, 42:D7−D17

14. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW: The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 2014, 42: D986–D992
15. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI: Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. Am J Hum Genet 2008, 82:100–112
16. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: A map of human genome variation from population-scale sequencing. Nature 2010, 467:1061–1073
17. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR: ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014, 42:D980–D985
18. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE: ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med 2008, 10:294–300
19. Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P: Assessing pathogenicity: overview of results from the IARC Unclassified Genetic Variants Working Group. Hum Mutat 2008, 29:1261–1264
20. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ: Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 2004, 75:535–544
21. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, et al: 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 2014, 46:107–115
22. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE: A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 2007, 81:873–883
23. Greaves M, Maley CC: Clonal evolution in cancer. Nature 2012, 481: 306–313
24. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C: Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012, 366:883–892
25. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, Olshen RA, Weyand CM, Boyd SD, Goronzy JJ: Diversity and clonal selection in the human T-cell repertoire. Proc Natl Acad Sci U S A 2014, 111: 13139–13144
26. Schumacher JA, Duncavage EJ, Mosbruger TL, Szankasi PM, Kelley TW: A comparison of deep sequencing of TCRG rearrangements vs traditional capillary electrophoresis for assessment of clonality in T-cell lymphoproliferative disorders. Am J Clin Pathol 2014, 141:348–359
27. Lee LA, Wang Y, Maiese R, Arvai KJ, Pan Q, Mehta K, Caruthers SM, Gersen SL, Billouin-Frazier S, Racke FK, Jones D: Assessment of mutation status in a large series of patients with suspected cytopenias with normal karyotypes and without increased blasts [abstract]. Blood 2014, 124:4606
28. Lacey JV Jr, Mutter GL, Nucci MR, Ronnett BM, Ioffe OB, Rush BB, Glass AG, Richesson DA, Chatterjee N, Langholz B, Sherman ME: Risk of subsequent endometrial carcinoma associated with endometrial intraepithelial neoplasia classification of endometrial biopsies. Cancer 2008, 113:2073–2081
29. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013, 368:2059–2074
30. Cervera N, Itzykson R, Coppin E, Prebet T, Murati A, Legall S, Vey N, Solary E, Birnbaum D, Gelsi-Boyer V: Gene mutations differently impact the prognosis of the myelodysplastic and myeloproliferative classes of chronic myelomonocytic leukemia. Am J Hematol 2014, 89:604–609
31. Patnaik MM, Padron E, Laborde RR, Lasho TL, Finke CM, Hanson CA, Hodnefield JM, Knudson RA, Ketterling RP, Al-Kali A, Pardanani A, Ali NA, Komrokji RS, Tefferi A: Mayo prognostic model for WHO-defined chronic myelomonocytic leukemia: ASXL1 and spliceosome component mutations and outcomes. Leukemia 2013, 27:1504–1510
32. Schnittger S, Bacher U, Eder C, Dicker F, Alpermann T, Grossmann V, Kohlmann A, Kern W, Haferlach C, Haferlach T: Molecular analyses of 15,542 patients with suspected BCR-ABL1-negative myeloproliferative disorders allow to develop a stepwise diagnostic workflow. Haematologica 2012, 97:1582–1585
33. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA: Integrated genomic characterization of endometrial carcinoma. Nature 2013, 497:67–73
34. Tothill RW, Li J, Mileshkin L, Doig K, Siganakis T, Cowin P, Fellowes A, Semple T, Fox S, Byron K, Kowalczyk A, Thomas D, Schofield P, Bowtell DD: Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. J Pathol 2013, 231:413–423
35. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellstrom-Lindberg E, Tefferi A, Bloomfield CD: The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 2009, 114:937–951
36. Fasan A, Alpermann T, Haferlach C, Grossmann V, Roller A, Kohlmann A, Eder C, Kern W, Haferlach T, Schnittger S: Frequency and prognostic impact of CEBPA proximal, distal and core promoter methylation in normal karyotype AML: a study on 623 cases. PLoS One 2013, 8:e54365
37. Jabbour E, Jones D, Kantarjian HM, O'Brien S, Tam C, Koller C, Burger JA, Borthakur G, Wierda WG, Cortes J: Long-term outcome of patients with chronic myeloid leukemia treated with second-generation tyrosine kinase inhibitors after imatinib failure is predicted by the in vitro sensitivity of BCR-ABL kinase domain mutations. Blood 2009, 114:2037–2043
38. Mok T, Yang JJ, Lam KC: Treating patients with EGFR-sensitizing mutations: first line or second line—is there a difference? J Clin Oncol 2013, 31:1081–1088
39. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA: Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011, 364:2507–2516
40. Puig N, Sarasquete ME, Balanzategui A, Martinez J, Paiva B, Garcia H, Fumero S, Jimenez C, Alcoceba M, Chillon MC, Sebastian E, Marin L, Montalban MA, Mateos MV, Oriol A, Palomera L, de la Rubia J, Vidriales MB, Blade J, Lahuerta JJ, Gonzalez M, Miguel JF, Garcia-Sanz R: Critical evaluation of ASO RQ-PCR for minimal residual disease evaluation in multiple myeloma: a comparative analysis with flow cytometry. Leukemia 2014, 28:391–397
41. Wu D, Sherwood A, Fromm JR, Winter SS, Dunsmore KP, Loh ML, Greisman HA, Sabath DE, Wood BL, Robins H: High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. Sci Transl Med 2012, 4:134ra63
42. Stead LF, Sutton KM, Taylor GR, Quirke P, Rabbitts P: Accurately identifying low-allelic fraction variants in single samples with next-generation sequencing: applications in tumor subclone resolution. Hum Mutat 2013, 34:1432–1438
43. Chen K, Meric-Bernstam F, Zhao H, Zhang Q, Ezzeddine N, Tang LY, Qi Y, Mao Y, Chen T, Chong Z, Zhou W, Zheng X, Johnson A, Aldape KD, Routbort MJ, Luthra R, Kopetz S, Davies MA, de Groot J, Moulder S, Vinod R, Farhangfar CJ, Shaw KM, Mendelsohn J,

Mills GB, Eterovic AK: Clinical actionability enhanced through deep targeted sequencing of solid tumors. Clin Chem 2015, 61:544–553
44. Lin CC, Hou HA, Chou WC, Kuo YY, Liu CY, Chen CY, Lai YJ, Tseng MH, Huang CF, Chiang YC, Lee FY, Liu MC, Liu CW, Tang JL, Yao M, Huang SY, Ko BS, Wu SJ, Tsay W, Chen YC, Tien HF: IDH mutations are closely associated with mutations of DNMT3A, ASXL1 and SRSF2 in patients with myelodysplastic syndromes and are stable during disease evolution. Am J Hematol 2014, 89:137–144
45. Daber R, Sukhadia S, Morrissette JJ: Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets. Cancer Genet 2013, 206:441–448
46. Spencer DH, Abel HJ, Lockwood CM, Payton JE, Szankasi P, Kelley TW, Kulkarni S, Pfeifer JD, Duncavage EJ: Detection of FLT3 internal tandem duplication in targeted, short-read-length, next-generation sequencing data. J Mol Diagn 2013, 15:81–93
47. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA: COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011, 39:D945–D950
48. Bell DW, Gore I, Okimoto RA, Godin-Heymann N, Sordella R, Mulloy R, Sharma SV, Brannigan BW, Mohapatra G, Settleman J, Haber DA: Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. Nat Genet 2005, 37:1315–1316
49. Metzeler KH, Maharry K, Radmacher MD, Mrozek K, Margeson D, Becker H, Curfman J, Holland KB, Schwind S, Whitman SP, Wu YZ, Blum W, Powell BL, Carter TH, Wetzler M, Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G, Bloomfield CD: TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol 2011, 29:1373–1381
50. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Sole F, Bennett JM, Bowen D, Fenaux P, Dreyfus F, Kantarjian H, Kuendgen A, Levis A, Malcovati L, Cazzola M, Cermak J, Fonatsch C, Le Beau MM, Slovak ML, Krieger O, Luebbert M, Maciejewski J, Magalhaes SM, Miyazaki Y, Pfeilstocker M, Sekeres M, Sperr WR, Stauder R, Tauro S, Valent P, Vallespi T, van de Loosdrecht AA, Germing U, Haase D: Revised international prognostic scoring system for myelodysplastic syndromes. Blood 2012, 120:2454–2465
51. Malcovati L, Hellstrom-Lindberg E, Bowen D, Ades L, Cermak J, Del CC, Della Porta MG, Fenaux P, Gattermann N, Germing U, Jansen JH, Mittelman M, Mufti G, Platzbecker U, Sanz GF, Selleslag D, Skov-Holm M, Stauder R, Symeonidis A, van de Loosdrecht AA, de WT, Cazzola M: Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. Blood 2013, 122:2943–2964
52. Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W: The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. Leukemia 2014, 28:485–496
53. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, Liu XS, Aravind L, Agarwal S, Maciejewski JP, Rao A: Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 2010, 468:839–843
54. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landén M, Höglund M, Lehmann S, Gabriel SB, Moran JL, Lander ES, Sullivan PF, Sklar P, Grönberg H, Hultman CM, McCarroll SA: Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 2014, 371:2477–2487
55. Gaidzik VI, Paschka P, Spath D, Habdank M, Kohne CH, Germing U, von Lilienfeld-Toal M, Held G, Horst HA, Haase D, Bentz M, Gotze K, Dohner H, Schlenk RF, Bullinger L, Dohner K: TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. J Clin Oncol 2012, 30:1350–1357
56. Smith AE, Mohamedali AM, Kulasekararaj A, Lim Z, Gaken J, Lea NC, Przychodzen B, Mian SA, Nasser EE, Shooter C, Westwood NB, Strupp C, Gattermann N, Maciejewski JP, Germing U, Mufti GJ: Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. Blood 2010, 116:3923–3932
57. Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, Shi YG, Zhu J, Wang P, Xu Y: Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell 2013, 155:1545–1555
58. Lin TL, Nagata Y, Kao HW, Sanada M, Okuno Y, Huang CF, Liang DC, Kuo MC, Lai CL, Lee EH, Shih YS, Tanaka H, Shiraishi Y, Chiba K, Lin TH, Wu JH, Miyano S, Ogawa S, Shih LY: Clonal leukemic evolution in myelodysplastic syndromes with TET2 and IDH1/2 mutations. Haematologica 2014, 99:28–36
59. Nibourel O, Kosmider O, Cheok M, Boissel N, Renneville A, Philippe N, Dombret H, Dreyfus F, Quesnel B, Geffroy S, Quentin S, Roche-Lestienne C, Cayuela JM, Roumier C, Fenaux P, Vainchenker W, Bernard OA, Soulier J, Fontenay M, Preudhomme C: Incidence and prognostic value of TET2 alterations in de novo acute myeloid leukemia achieving complete remission. Blood 2010, 116:1132–1135
60. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB: In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat 2008, 29:1327–1336
61. Chan PA, Duraisamy S, Miller PJ, Newell JA, McBride C, Bond JP, Raevaara T, Ollila S, Nyström M, Grimm AJ, Christodoulou J, Oetting WS, Greenblatt MS: Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). Hum Mutat 2007, 28:683–693
62. Gemovic B, Perovic V, Glisic S, Veljkovic N: Feature-based classification of amino acid substitutions outside conserved functional protein domains. Scientific World Journal 2013, 2013:948617
63. Gonzalez-Perez A, Lopez-Bigas N: Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 2011, 88:440–449
64. Kroeze LI, Aslanyan MG, van RA, Koorenhof-Scheele TN, Massop M, Carell T, Boezeman JB, Marie JP, Halkes CJ, de WT, Huls G, Suciu S, Wevers RA, van der Reijden BA, Jansen JH: Characterization of acute myeloid leukemia based on levels of global hydroxymethylation. Blood 2014, 124:1110–1118
65. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al: Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013, 499:214–218
66. Weissmann S, Alpermann T, Grossmann V, Kowarsch A, Nadarajah N, Eder C, Dicker F, Fasan A, Haferlach C, Haferlach T, Kern W, Schnittger S, Kohlmann A: Landscape of TET2 mutations in acute myeloid leukemia. Leukemia 2012, 26:934–942
67. Kutny MA, Alonzo TA, Gerbing RB, Ho PA, Geraghty D, Lange B, Heerema NA, Meshinchi S: TET2 SNP rs2454206 (I1762V) correlates with improved survival in pediatric acute myelogenous leukemia, a report from the Children's Oncology Group [abstract]. Blood 2010, 116:418a
68. Lindor NM, Guidugli L, Wang X, Vallee MP, Monteiro AN, Tavtigian S, Goldgar DE, Couch FJ: A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). Hum Mutat 2012, 33:8–21
69. McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummer R, Ribas A, Hogg D, Hamid O, Ascierto PA, Garbe C, Testori A, Maio M, Lorigan P, Lebbe C, Jouary T, Schadendorf D, O'Day SJ, Kirkwood JM, Eggermont AM, Dreno B, Sosman JA, Flaherty KT, Yin M, Caro I, Cheng S, Trunzer K, Hauschild A: Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended

follow-up of a phase 3, randomised, open-label study. Lancet Oncol 2014, 15:323−332
70. Landsman-Blumberg PB, Carter GC, Johnson BH, Sedgley R, Nicol SJ, Li L, Shankaran V: Metastatic colorectal cancer treatment patterns according to kirsten rat sarcoma viral oncogene homolog genotype in U.S. community-based oncology practices. Clin Colorectal Cancer 2014, 13:178−184
71. Qu K, Pan Q, Zhang X, Rodriguez L, Zhang K, Li H, Ho A, Sanders H, Sferruzza A, Cheng SM, Nguyen D, Jones D, Waldman F: Detection of BRAF V600 mutations in metastatic melanoma: comparison of the Cobas 4800 and Sanger sequencing assays. J Mol Diagn 2013, 15:790−795
72. Ding D, Yu Y, Li Z, Niu X, Lu S: The predictive role of pretreatment epidermal growth factor receptor T790M mutation on the progression-free survival of tyrosine-kinase inhibitor-treated non-small cell lung cancer patients: a meta-analysis. Onco Targets Ther 2014, 7:387−393
73. Bradeen HA, Eide CA, O'Hare T, Johnson KJ, Willis SG, Lee FY, Druker BJ, Deininger MW: Comparison of imatinib mesylate, dasatinib (BMS-354825), and nilotinib (AMN107) in an N-ethyl-N-nitrosourea (ENU)-based mutagenesis screen: high efficacy of drug combinations. Blood 2006, 108:2332−2338
74. Cottrell CE, Al-Kateb H, Bredemeyer AJ, Duncavage EJ, Spencer DH, Abel HJ, Lockwood CM, Hagemann IS, O'Guin SM, Burcea LC, Sawyer CS, Oschwald DM, Stratman JL, Sher DA, Johnson MR, Brown JT, Cliften PF, George B, McIntosh LD, Shrivastava S, Nguyen TT, Payton JE, Watson MA, Crosby SD, Head RD, Mitra RD, Nagarajan R, Kulkarni S, Seibert K, Virgin HW, Milbrandt J, Pfeifer JD: Validation of a next-generation sequencing assay for clinical molecular oncology. J Mol Diagn 2014, 16:89−105
75. Jung H, Bleazard T, Lee J, Hong D: Systematic investigation of cancer-associated somatic point mutations in SNP databases. Nat Biotechnol 2013, 31:787−789