BMC Cancer

Portrait of a cancer: mutational signature analyses for cancer diagnostics

Arne Van Hoeck[1*] , Niels H. Tjoonk[1,2], Ruben van Boxtel[2] and Edwin Cuppen[1,3]

Abstract

Background: In the past decade, systematic and comprehensive analyses of cancer genomes have identified cancer driver genes and revealed unprecedented insight into the molecular mechanisms underlying the initiation and progression of cancer. These studies illustrate that although every cancer has a unique genetic make-up, there are only a limited number of mechanisms that shape the mutational landscapes of cancer genomes, as reflected by characteristic computationally-derived mutational signatures. Importantly, the molecular mechanisms underlying specific signatures can now be dissected and coupled to treatment strategies. Systematic characterization of mutational signatures in a cancer patient's genome may thus be a promising new tool for molecular tumor diagnosis and classification.

Results: In this review, we describe the status of mutational signature analysis in cancer genomes and discuss the opportunities and relevance, as well as future challenges, for further implementation of mutational signatures in clinical tumor diagnostics and therapy guidance.

Conclusions: Scientific studies have illustrated the potential of mutational signature analysis in cancer research. As such, we believe that the implementation of mutational signature analysis within the diagnostic workflow will improve cancer diagnosis in the future.

Keywords: Mutational signature, Cancer diagnosis, Cancer biomarkers, Cancer genomics, Molecular medicine, Whole genome sequencing

Background

Historically, cancer diagnostic and treatment decisions were predominantly based on tumor morphology, clinical symptoms, and the cancer site of origin. In the past decade, systematic analyses of cancer genomes have changed this paradigm [1], and the term 'cancer' now encompasses more than a hundred different diseases differentiated on the basis of varying combinations of cancer gene mutations [2, 3]. This development, together with the emergence of molecularly targeted drugs, resulted in an increase in molecular testing to support decision making in cancer diagnostics and treatment.

Thus far, the development of cancer diagnostics has mainly focused on identifying driver mutations that provide growth advantages to cancer cells and thereby

promote tumorigenesis [2]. Genetic testing for driver genes can identify the biological characteristics of tumors. These genes can also act as direct targets for effective treatment. The rapidly growing number of drugs directly targeting proteins encoded by mutated driver genes has fueled the development of assays for the accurate detection of mutations for cancer diagnosis [4].

Although this knowledge has contributed significantly to drug development and improved cancer care, a substantial portion of patients do not benefit from this strategy because of poor response rates to targeted drugs and a lack of adequate biomarkers. Therefore, cancer diagnostics require better molecular characterization of tumors, as well as reliable biomarkers for patient stratification. Next-generation sequencing (NGS) technologies have emerged as an important tool to fulfill this unmet need. The capacity of NGS to analyze large panels of genes, up to complete cancer genomes, has enabled the generation of comprehensive catalogues of somatic mutations in cancer patients [5–7]. However, only a very small fraction of

* Correspondence: A.VanHoeck@umcutrecht.nl
[1]Center for Molecular Medicine and Oncode Institute, University Medical Centre Utrecht, Heidelberglaan 100, 3584CX Utrecht, The Netherlands
Full list of author information is available at the end of the article

Hoeck *et al. BMC Cancer* (2019) 19:457

Page 2 of 14

the identified variants are tumor drivers or actionable bio-markers. The vast majority of somatic mutations in a can-cer genome are passenger mutations, which are not believed to be involved in cancer development. Neverthe-less, it has recently been shown that these alterations can be used to provide insight into the history of the tumor and identify mutational processes that have occurred be-fore and during tumorigenesis [8]. Somatic mutations can originate from exogenous factors, such as environmental carcinogens or UV radiation, or endogenous processes, such as normal mutational decay due to spontaneous de-amination of methylated nucleotides, base misincorpora-tion by error-prone polymerases, and unrepaired or incorrectly repaired DNA damage due to impaired DNA damage response (DDR) gene function (reviewed by Hel-leday et al. [9]). Interestingly, each of these leave a charac-teristic pattern of mutations, which have been dubbed 'mutational signatures' [8]. For instance, cells defective in homologous recombination repair (HRR) machinery or non-dividing cells must rely on alternatives to repair DNA breaks, such as non-homologous end-joining and alterna-tive end-joining to repair double-stranded DNA breaks [10]. These repair processes are not error free and leave a characteristic mutational pattern, which has been shown to be useful for the identification of tumors deficient in HRR [11, 12]. Mutational signatures can therefore reflect the presence or absence of cellular processes in cancer cells. Because multiple endogenous or exogenous muta-tional forces can operate simultaneously or successively on the genome during a cell's life span, the mutational catalogue of a cancer genome harbors a mixture of signa-tures shaped by different mutational processes. Some of these mutational processes are active continuously throughout the lifetime of the cancer cell (clock signa-tures) [13], whereas others are active periodically, some of which are influenced by the patient's lifestyle [14].

It has recently been shown that mutational signatures can be biomarkers for specific characteristics of a cancer [8, 15]. As such, they bear potential clinical value as pre-dictors of the therapy response in cancer [11]. An import-ant prerequisite for mutational signature analysis is the availability of genome-wide mutational data across many independent cancers. As the cost of whole-genome se-quencing decreases and the amount of available cancer mutation data grows, it is timely to consider mutational signature analysis a novel opportunity for biomarker dis-covery, tumor diagnostics, and treatment guidance.

Signatures reveal mutation etiology
The first mutational signatures introduced were base substitutions. For these mutation types, a signature is characterized by the specific base change and its direct 5′ and 3′ flanking base. Because there are six classes of base substitution and 16 possible sequence contexts,

there are 96 distinguishable trinucleotide changes. Therefore, mutational signatures can be distilled from large cohorts of sequenced cancer patients by a compu-tational framework that attempts to decompose distin-guishable recurrent patterns from the cohort's 96-mutation matrix. Ultimately, each pattern represents the relative proportion of each trinucleotide mutation, which reflects a mutational signature. More theoretical details about the framework can be found in Alexandrov et al. [16], and Serena et al. [17] provides a chronological overview on mutational signature analysis in cancer.

Although mutational signatures are a relatively recent concept in cancer biology, the idea of linking mutational processes with mutational patterns is not new. The first studies linking specific mutation characteristics to vari-ous environmental mutagens, such as UV-radiation [18], smoking [19], and aristolochic acid [20], were focused on single cancer genes that were recurrently mutated in a wide range of cancers, such as *TP53* and *BRAF*. These studies provided the first evidence that mutational pro-cesses can leave characteristic patterns in the DNA that are visible and analyzable in tumor samples via the de-tection of distinct signatures [21]. In 2013, Stratton and his team introduced a computational framework that used nonnegative matrix factorization (NMF) to recognize multiple base substitution patterns in human cancers [15, 22]. Moreover, some of these patterns corre-lated with known mutagenic processes, indicating that this mathematical concept can extract biologically rele-vant information to unravel mechanisms underlying tumorigenesis [16]. Since this seminal study by Stratton's group, the field of mutational signature analysis has grown rapidly in cancer biology. Currently, there are 30 different reference signatures described in primary can-cer that are categorized in the COSMIC database (http://cancer.sanger.ac.uk/cosmic/signatures) [22]. However, additional signatures continue to be identified by various research groups [23–26], and methods to characterize cancer genomes in a similar way based on indels, structural variants, and copy-number changes are currently under development [27].

Comparing these signatures with the scientific litera-ture, as well as statistically associating them with patient phenotypes, provided the first mechanistic insights into the etiology of a number of mutational processes. Mech-anisms underpinning mutational signatures have been suggested for roughly half of the 30 COSMIC signatures. The establishment of large pan-cancer genomic datasets, such as The Cancer Genome Atlas (TCGA) [28], Wel-come Trust Sanger Institute's Cancer Genome Project [29] and the International Cancer Genome Consortium (ICGC) [1], were vital for these analyses. By doing so, exogenous processes (e.g., tobacco smoking and UV-exposure) and endogenous processes (e.g., APOBEC

Hoeck *et al. BMC Cancer*   (2019) 19:457

Page 3 of 14

overactivity, deficiency in double strand break repair, and polymerase slippage) could be attributed to specific signatures. However, obtaining evidence that the proposed etiology of a signatures is a specific mutational process, based solely on data derived from cancer patients, is not straightforward. It is complicated by the lack of complete catalogues of true pathogenic driver variants and missing information on the environmental exposure history of the patient cohort. Additional complexities can be found in the heterogeneous landscape of mutational processes that is typically identified in individual cancers. Furthermore, the detected somatic mutations are the result of a balance between mutation-inducing and DNA repair processes, which are not fully independent, and mechanisms may vary between tissues. Therefore, more controlled experimental approaches are needed to determine the origin of a signature. We recently showed that the application of CRISPR-Cas9 technology in human colon organoids to delete key genes involved in specific DNA repair pathways, followed by genome-wide characterization of the resulting mutation patterns, is a powerful approach because it can link the observed signatures of the accumulated mutations directly to the biological functionality of the inactivated gene [30].

Diagnostic mutational signatures

Currently, the most notable advances in mutational signature analysis-based diagnosis are in the field of breast cancer. Tumors with mutations in *BRCA1/2* are defective in the HRR process. These tumors show promising responses to treatment with a PARP inhibitor (olaparib), a drug that decreases the DDR in cancer cells to a fatally low level [31–33]. DNA-damaging agents that directly induce double strand breaks, such as chemotherapy based on platinum salts, prove therapeutically efficient in these cancers as well [34–36]. Recently, a model that can accurately predict HRR deficiency (HRDetect) was developed for breast cancers [11]. This computational tool uses HRR-deficiency features from the complete mutation catalogue of base substitutions, indels, and structural rearrangements. The use of this tool revealed that microhomology-mediated indels, two COSMIC signatures (further referred to as CS) and two rearrangement signatures (further referred to as RS) correlated with HRR deficiency (Fig. 1). By accounting for their mutational contribution, HRDetect could predict BRCA-ness (i.e., a *BRCA1/2*-associated phenotype) with a sensitivity of almost 100%, which is an improvement on the sensitivity obtained by more traditional copy number based tests (~ 60%) [37] and functional assays of HRR deficiency (~ 80%) [38]. HRDetect identified 44 cancers that carried a germline or a somatic *BRCA1/2* variant in a cohort of 560 breast cancer patients and, interestingly,

in 47 cancers demonstrating BRCAness in which no pathogenic variant in *BRCA1/2* was detected. The latter category can possibly be explained by the epigenetic inactivation of *BRCA1/2* or the inactivation of other components involved in HRR.

The HRDetect tool demonstrates that signature analysis can be deployed to successfully identify BRCAness in patients without the need for prior knowledge of *BRCA* mutations. Polak et al. [12] found similar results in a different breast cancer cohort, and pointed out that cancers carrying a somatic event in *BRCA1* ($n = 36$, cohort size = 995) or *BRCA2* ($n = 34$, 995) had a stronger contribution from CS-3. Interestingly, cancers that showed epigenetic silencing of *BRCA1* ($n = 32$, 995) or *RAD51C* ($n = 23$, 995), or that carried germline *PALB2* ($n = 3$, 995) or *RAD51C* ($n = 1$, 995) mutations, also displayed an increased contribution from CS-3. Because epigenetic modifications cannot be directly verified by traditional diagnostic methods, identifying mutational signatures associated with HRR defects can increase the number of patients who would benefit from treatment with PARPi and platinum-based drugs [11]. Recently, we validated this strategy in breast cancer organoids by subjecting organoids derived from a patient who displayed a high contribution from CS-3 mutations to two different PARPi drugs [39]. These organoids were sensitive to PARPi, whereas breast cancer organoids negative for CS-3 did not show any response, illustrating the principle that CS-3 can act as a useful marker for PARPi sensitivity in cancer. A recent retrospective study computed HRDetect scores for 93 advanced breast cancer patients, 33 of which were treated with platinum chemotherapy [34]. All patients scoring high for HRR deficiency showed a significantly association with clinical improvement on platinum-based therapy. These findings provide evidence for the use of mutational signatures as sensitive biomarkers for HRR defects, and can inspire the design of therapeutic trials.

Moreover, mutational signature analysis to detect HRR deficiency could be applied to many different cancer types beyond breast cancer. Germline mutations in *BRCA1/2* have long been known to affect the risk of ovarian cancer [40] and pancreatic cancer [41]. Biomarkers for HRR deficiency were found in 24 additional cancer classes or cancer-associated syndromes [8, 42–44]. These findings suggest that HRR deficiency and the associated therapeutic benefits may apply to a greater number of patients than is currently appreciated. Indeed, in a study on pancreatic cancer, all patients that responded to platinum-based chemotherapy harbored the BRCA-associated CS-3 [45]. These examples indicate that an effective response to specific anti-cancer drugs is more dependent on specific functional defects in a tumor than by the organ in which this tumor is located.

Hoeck *et al. BMC Cancer* (2019) 19:457

Page 4 of 14

A Mutational signatures useful in analysis

A						B Underlying mutational process	C Relevant genes	D Predisposition syndrome	E Proposed therapy choice
CS-3	CS-8	MH-indels	RS-3	RS-5	HRD index	Homologous Recombination Repair Deficiency	BRCA1, BRCA2, RAD51C, PALB2	Hereditary Breast and Ovarian Cancer Syndrome	PARP inhibition[32-34], Platinum-based chemotherapy[35-37]
CS-6	CS-15	CS-20	CS-26	STR-indels		Mismatch Repair Deficiency	MLH1, MSH2, MSH6, PMS1, PMS2	Lynch, CMMRD, BMMR-D, HNPCC	PD1-immunotherapy[48-49,52]
CS-5	CS-8	TSB-sign				Nucleotide Excision Repair Deficiency	ERCC1, ERCC2, XPC	Xeroderma Pigmentosum	Cisplatin[63-65]
CS-18	CS-30	TSB-sign	C>A*	G>T*	C>T*	Base excision Repair Deficiency	MUTYH,OGG1	MAP	
							NTHL1, SMUG1	NAP	
CS-10	STR-indels					Deficient DNA polymerase proofreading activity	POLE, POLD1	PPAP	PD1-immunotherapy[48-49,52]
?						Non-Homologous End Joining Deficiency		Nijmegen Breakage Syndrome	
CS-2	CS-13	Kataegis				APOBEC Over-activity	APOBEC1, APOBEC3A, APOBEC3B		Tamoxifen Resistance[70,71]

Fig. 1 Mutational processes linked to treatment selection via mutational signatures. Mutational signatures in tumor genomes can reflect the activity of specific mutational processes and thereby provide support for therapy selection. Different types of mutational signatures (**a**) can be considered: base substitution signatures (orange), indel signatures (green), rearrangement signatures (yellow), geographically localized mutational phenomena (blue) or other signatures characterized by copy-number variations (grey). Diagnostic interpretation of characteristic signatures can contribute to therapy choice (**e**) and include (green) or exclude (red) patients from a treatment. Actionable pathways that can be identified by mutational signatures (**a**) mainly include DNA repair defects (**b**), which was confirmed by the presence of pathogenic mutations in the indicated genes in these pathways (**c**). The prevalence of germline pathogenic mutations in these genes is typically linked to a cancer predisposition syndrome (**d**). Abbreviations: CS-[number], COSMIC signature; RS-[number], rearrangement signature; MH-indels, indels at microhomologies; STR-indels, short tandem repeat-mediated indels; TSB sigs, signatures showing transcriptional strand bias. APOBEC, apolipoprotein B DNA-editing complex; MAP, MUTYH-associated polyposis; NAP, NTHL1-associated polyposis; PARP, poly(ADP-ribose) polymerase; PPAP, polymerase proofreading associated polyposis. * defects in base excision repair have been associated with these characteristic substitutions

Nevertheless, the efficacy of HRDetect in selecting patients of all cancer types for PARPi, platinum-based, and/or immune-based therapy needs testing in (pre)clinical trials.

Similar tactics could be employed for other mutational process signatures as well. DNA mismatch repair (MMR) corrects stochastic errors by polymerases that arise during DNA replication [46]. A deficiency in MMR and DNA proofreading results in increased mutational load of base substitutions and instability at tandem repeats of short nucleotide sequences (a feature called microsatellite instability [MSI]) [47] (Fig. 1). Colorectal cancers with MMR deficiency are sensitive to pembrolizumab [48] and nivolumab [49], which are both inhibitors of the programmed death 1 (PD1) immune checkpoint. In 2015, the Consensus Molecular Subtypes (CMS) Consortium subcategorized all hypermutated MSI cancers in one CMS group (CMS1, 14% of colorectal cancers) based on gene expression data. Mutational signature analysis demonstrated that MSI colorectal cancers leave specific mutational signatures (CS-6, CS-15, CS-10, CS-20, and CS-26) [8, 9], which can be used to identify MMR deficiency in cancers [26, 30]. Recently, we validated the association between MMR deficiency and a CS-20-like signature in colon organoids that lack the essential MMR gene *MLH1* [30]. These organoids were exclusively characterized by this base substitution signature accompanied by small indels (< 3 bp) within a tandem repeat context (Fig. 1). These mutation characteristics could be used to identify colon cancer patients with MMR deficiency even when that deficiency is caused by epigenetic mechanisms such as the well-studied *MLH1* promotor methylation. Although MRR-deficient cancers dominate in colorectal cancers [50], signature analysis revealed MMR-deficient pancreatic cancer as well (*n* = 3, 180) [51]. Thus, as in the case of HRR deficiency, signature analysis might be a convenient approach to simultaneously screen for MMR deficiency to identify patients who would benefit from immunotherapy, regardless of the cancer's tissue of origin [52]. Indeed, in a follow up study, Le et al. showed that PD1 inhibition is not just successful in treating colon cancer with MSI but also in treating 11 other cancer types with MMR-deficiency [53].

Base excision repair (BER) is a third category of DNA repair that could potentially be discerned by mutational signatures. Defects in BER components *SMUG1*, *OGG1*, and *NTHL1* result in higher rates of C > A transversions (*SMUG1* and *OGG1*) [54, 55] and C > T transitions (*SMUG1* and *NTHL1*) [56, 57]. These findings indicate that the failure of BER processes might also leave specific predictive marks. Indeed, using CRISPR/Cas9-mediated knockout of *NTHL1* in colon organoids, we have shown that *NTHL1* deficiency results in increased mutations, which can be attributed to CS-30 [30]. This signature had been identified in only a single cancer patient within a breast cancer cohort [22]. Upon examining the germline of this patient, we identified a heterozygous mutation causing a premature stop codon in *NTHL1*, with loss of heterozygosity in the tumor. Mutations in *MUTYH*, a BER- and nucleotide excision repair (NER)--associated protein, are specifically associated with CS-18 [58] and a CS-18-like signature [26, 59]. Because BER and NER can both be coupled to transcription [60, 61], more specific mutational signatures could possibly be dissected when such genomic features are taken into account (including CS-4, CS-5, CS-8, CS-12, CS-16, and CS-22 – see Fig. 1) [8]. For example, a specific mutational signature that closely resembles CS-5 has been associated with defects in *ERCC2*, a core protein of the NER pathway [62]. Importantly, this signature was significantly increased in responders to cisplatin compared to non-responders, and other studies have also confirmed a positive response to cisplatin in NER-deficient patients [63–65]. However, the studies of CS-5 also illustrate one of the limitations of the use of mutational signatures. It is now considered that this siganture represents a universal ageing signature, as does CS-1 [13, 30], since both signatures have been observed in healthy cells. CS-5 therefore has little diagnostic value, but it remains to be shown whether quantitative analyses reveal a robust association of NER deficiency with increased levels of CS-5 mutations. Furthermore, not all NER-deficient tumors show the same signature contribution, suggesting that distinct mutational processes related to NER deficiency might be active. Indeed, recent findings from our laboratory indicate that deficiency in global genome NER results in a tissue-specific increase in mutations, which can be attributed to CS-8 [66].
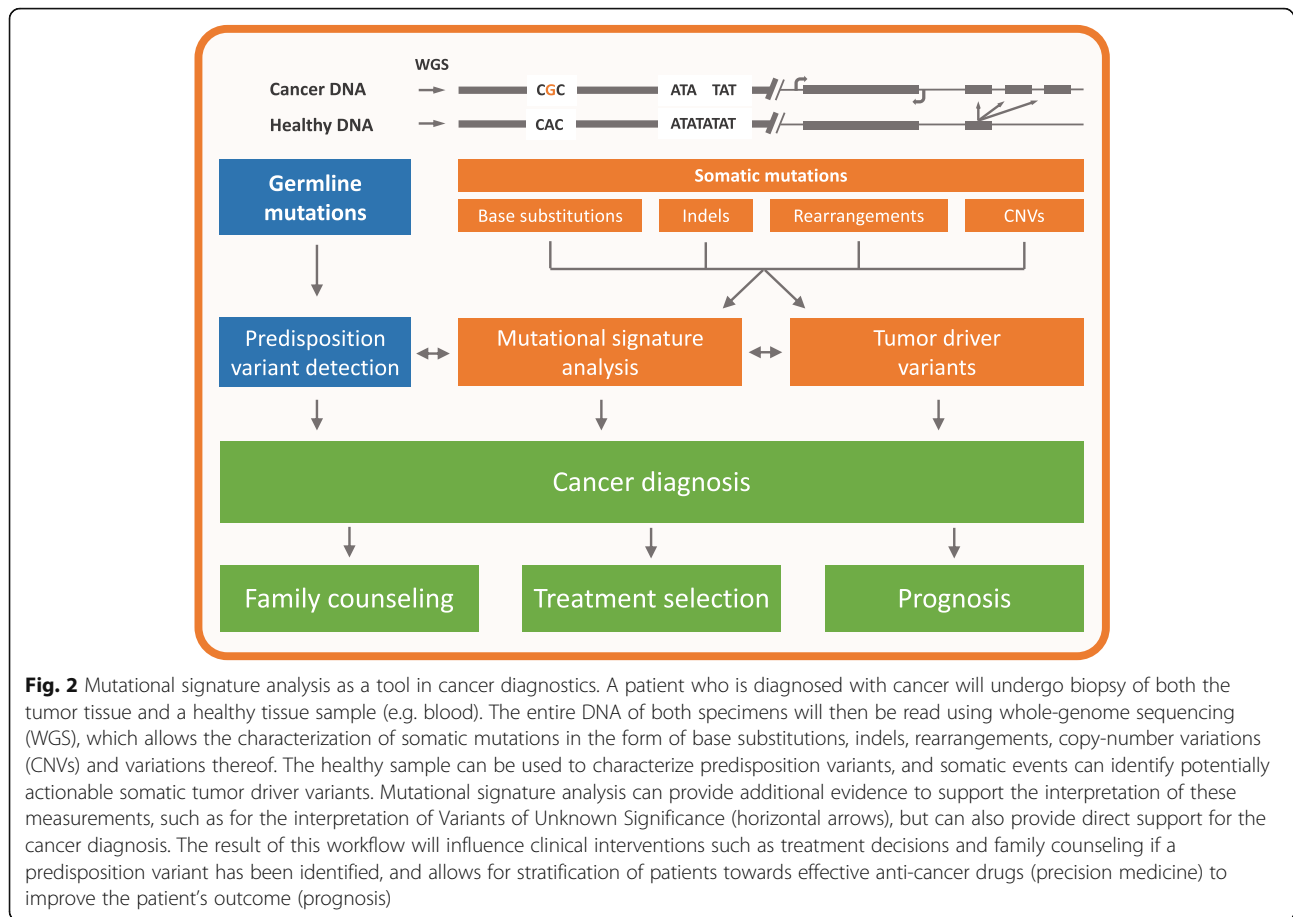
In addition to DNA repair deficiencies, other cellular processes can leave informative signatures in tumors. Activation of the RNA-editing enzyme APOBEC constitutes part of the cellular immune response to viruses and retrotransposons, but overactivity of APOBEC is a driving force of somatic hypermutation [67]. This implies that tumors with APOBEC overactivity could be treated by lethal mutagenesis, which consists of administration of drugs stimulating mutation rates past a lethal threshold, thereby stimulating programmed cell death [68]. APOBEC enzymes have also been proposed to drive cancer evolution, heterogeneity, and therapy resistence [69]. APOBEC overactivity has been shown to promote drug resistance to the cancer drug Tamoxifen [70, 71], perhaps due to APOBEC-driven intratumor heterogeneity. The APOBEC-associated signatures CS-2 and CS-8, as well as an associated phenomenon of clustered mutagenesis called kataegis (Fig. 1), have been found in more than half of the investigated cancer types [6]. Additionally, later studies found these signatures in in a range of cancer types [24, 72–74] and directly linked them to an *APOBEC3A/3B* germline deletion allele in breast cancer [75]. Detection of APOBEC overactivity could therefore be useful in a wide range of cancer types. Moreover, mutational signature analysis allows discrimination between the signatures of different APOBEC-subtypes [76]; the APOBEC3B subtype could be further subdivided with clustered mutational signatures [77], which means even more specific targeting could be possible. For example, APOBEC stimulators might be used to stimulate lethal mutagenesis.

Stratification of cancer patients

In addition to using mutational signatures as a genomic biomarker for targeted therapeutics, mutational signature analysis presents possibilities in the stratification of patients (Fig. 2). For instance, breast cancer is among the most common types of cancer worldwide, with an estimated incidence of 1.7 million cases in 2012 [78]. Around 5–10% of all breast cancers are attributed to somatic or germline mutations in the genes *BRCA1* and *BRCA2* [79]. However, HRR deficiency is currently not an intrinsic subclass in breast cancer diagnostics, although this cohort may have a better prognosis when treated with specific drugs. A few recent studies have applied mutational signature analysis to identify which patients are most likely to respond to certain therapies, including studies of patients with esophageal adenocarcinoma (EAC) [43], pancreatic ductal adenocarcinoma (PDAC) [44], oral squamous cell carcinoma (OSCC) [80], gastric cancer [25], and prostate cancer [81].

EAC is an illustrative example. Highly heterogeneous mutational landscapes and a current lack of efficient stratification methods has led to the generally poor performance of targeted therapeutic approaches [43, 82]. However, in a cohort of 129 EAC patients, Secrier et al. [43] were able to define each patient's tumor by its dominant mutational signature and performed hierarchical clustering to stratify tumors into three subgroups with distinct etiologies. The first subgroup exhibited faults in the HRR pathway and was characterized by CS-3, and could therefore benefit from PARP inhibitors or platinum-based chemotherapy. The largest subgroup

Fig. 2 Mutational signature analysis as a tool in cancer diagnostics. A patient who is diagnosed with cancer will undergo biopsy of both the tumor tissue and a healthy tissue sample (e.g. blood). The entire DNA of both specimens will then be read using whole-genome sequencing (WGS), which allows the characterization of somatic mutations in the form of base substitutions, indels, rearrangements, copy-number variations (CNVs) and variations thereof. The healthy sample can be used to characterize predisposition variants, and somatic events can identify potentially actionable somatic tumor driver variants. Mutational signature analysis can provide additional evidence to support the interpretation of these measurements, such as for the interpretation of Variants of Unknown Significance (horizontal arrows), but can also provide direct support for the cancer diagnosis. The result of this workflow will influence clinical interventions such as treatment decisions and family counseling if a predisposition variant has been identified, and allows for stratification of patients towards effective anti-cancer drugs (precision medicine) to improve the patient's outcome (prognosis)

predominantly showed CS-17, a signature that does not yet have a defined etiology but could be related to gastroesophageal reflux [83]. In this subgroup of cancers, an increased response to WEE1/CHK1 inhibitors was observed. In addition, CS-17 has been shown to correlate with high neoantigen loads, which could implicate these patients for immunotherapy [84–86]. The final subgroup predominantly showed the signatures CS-1, which is age-related, and CS-18, which has no consensus etiology as yet but has been suggested to be associated with damage from reactive oxygen species (ROS) [26, 59]. Although Secrier et al. suggested traditional chemotherapy for these patients, the clinical meaningfulness of this subgroup is questionable, because there is no other obvious treatment alternative for these patients at this time. Uncovering the mechanisms underlying CS-18 and CS-1 will possibly energize the search for therapeutic potential in this subgroup. Molecular stratification of cancer patients based on mutational signatures is used in a growing number of studies, although in variable forms. Whereas EACs and gastric cancers were classified using predominantly mutational signature analysis [25, 43], PDACs and OSCCs were stratified using mutational signatures as part of an integrated genomics approach

[44, 80]. However, other tumor characteristics must often contribute to a comprehensive tumor diagnosis and treatment decision, because not all therapies are directly related to the mutational processes driving cancer. Nevertheless, mutational signatures already provide relevant information for treatment selection in at least some subgroups. In addition, evaluation of mutational signatures is an interesting approach that could be explored for the stratification of patients in clinical trials.

Revealing cancer predisposition
The majority of cancers are believed to result from somatic mutations [2]. Nevertheless, up to 10% of the cases can be attributed to inherited variants present in the patient's germline [87]. Exome sequencing studies in the last decade have revealed many new predisposition gene candidates, and whole-genome sequencing (WGS) pan-cancer studies will likely unravel new predisposition genes in the future, such as non-coding driver variants [88]. Mutational signature analysis could potentially be applied as a powerful screening tool to uncover new pathogenic inherited mutations affecting mutation accumulation, and as a validation method to accurately classify variants of uncertain significance (VUS) as either

pathogenic or benign (Fig. 2). For instance, we have identified a germline *NTHL1* variant in a breast cancer patient by screening for CS-30 conribution [30]. Polak et al. revealed that nearly all samples showing a pathogenic *BRCA1/2* germline variant and loss of the intact allele were positive for CS-3 in the TCGA breast cancer co-hort, and accurately classified 12 *BRCA1/2* VUSs [12]. The integration of indel and rearrangement signatures can even segregate *BRCA1* deficient tumors from *BRCA2* mutants [11]. It is worth mentioning that not all heritable breast cancers harbor germline variants solely in *BRCA1/2*, indicating that predisposing variants in other genes likely exist and contribute to hereditary breast cancer via altered mutation accumulation.

A more advanced approach could incorporate the evo-lutionary dynamics of the signatures to identify early-onset signatures which, together with true driver detection, can be used to trace predisposition variants from tumor-only sequencing data [89]. Such an ap-proach has been tested in 15 ultrahypermutated cancer patients (> 100 mutations per Mb) and in each individ-ual, a germline MMR mutation was found. Moreover, this analysis was performed on panel sequencing data, which covers a sufficient number of nucleotides to iden-tify early-onset signatures from highly mutated cancer types but is likely not adequate for less mutated cancer types. However, this strategy can be implemented in a whole-genome/−exome framework to predict predispos-ition variants in other cancer types.

The application of mutational signature analysis to re-veal cancer predisposition could be an important step forward in familial cancer diagnosis. For instance, many colorectal cancer patients harbor mutated predisposition genes that can be classified into distinct colorectal can-cer subtypes including polymerase proofreading associ-ated polyposis (PPAP), *MUTYH*-associated polyposis (MAP), *NTHL1*-associated polyposis (NAP), and Lynch syndrome. These subtypes are pathologically very similar and therefore difficult to identify, requiring extensive multifactorial testing [90, 91]. However, PPAP has been associated with a distinct mutational signature, CS-10 [8]; MAP with two signatures, one CS-18-like [58] and a simi-lar signature currently named signature 36 [26]; and NAP with CS-30 [30, 57]. The clinical value of detecting these predispositions is shown in Fig. 1. In addition, Lynch syn-drome can be identified using the MMR-associated signa-ture CS-6 [8] and indel signatures [30]. Indeed, a study aiming to detect Lynch syndrome used the aforemen-tioned two Lynch syndrome-associated signatures and the PPAP-associated signature CS-10 to distinguish these two groups of patients [89]. However, for most of these syn-dromes, more research is required to validate the signa-tures. Additional studies of larger, selected cohorts can help unravel which syndromes are linked to which

signatures. In addition, it is important to study whether other predisposition syndromes, not functionally linked with DNA repair deficiency, can be associated with a spe-cific mutational pattern. These studies might best focus on hereditary cancer syndromes that are currently difficult to identify with targeted gene panels, such as Cowden syn-drome [92]. Furthermore, additional studies are necessary to evaluate the efficacy of mutational signature analysis in identifying different hereditary cancer types, particularly because different syndromes may converge on the same signature and be indistinguishable. Nevertheless, the as-signment of germline mutations in cancer patients has several important clinical implications, because these vari-ants can serve as sentinels for identifying families with high risk for cancer development. Family members carry-ing pathogenic germline variants could be encouraged to obtain genetic counseling, take preventive measures, or enter increased surveillance programs (Fig. 2).

Identifying tumor tissue of origin
Roughly 3% of all new cancer cases are diagnosed as a cancer of unknown primary (CUP) [93]. Furthermore, substantial uncertainty about the tissue of origin re-mains, especially when the cancer is metastatic or poorly differentiated; this complicates treatment because most targeted drugs are tumor type-specific. Mutational se-quencing data could support histopathological examin-ation in identifying the cancer site of origin. Comprehensive mutational signature analyses have shown that tumor types leave distinctive patterns of somatic mutations. For example, CS-12 and CS-16 are so far exclusively associated with liver cancer [94], and ovarian cancer typically harbors a high number of struc-tural variants [95]. Such tissue-specific patterns, or a combination thereof, could be exploited to accurately decipher the primary tissue type. The ICOMS [96] (in-ferring cancer origins from mutation spectra) tool and TumorTracer [97] are two examples of well-trained clas-sifiers that utilize TCGA and COSMIC data to infer the origin of distinct primary tumor sites. Although these tools deliver performance scores that may be accurate enough to aid in the clinical diagnosis of CUPs, the use of pan-cancer WGS data and advanced signature extrac-tion methods will likely lead to more accurate ap-proaches [98].

Existing challenges
Thus far, we have discussed the current state and poten-tial diagnostic value of mutational signature analysis, as well as applications for the detection of germline predis-position mutations and the determination of organ of origin for CUPs. However, clinical integration of such detection requires critical examination and further re-finement of these signatures, and some obvious

Hoeck *et al. BMC Cancer* (2019) 19:457

Page 8 of 14

weaknesses and limitations must be addressed. First, the current 30 COSMIC signatures are derived from a mix of whole-exome sequencing (WES) and WGS data (10,952 whole exomes and 1048 whole genomes). This has resulted in discrepancies between WES- and WGS-derived signatures; for example, certain processes specifically act on coding or non-coding elements, such as transcription-coupled repair. This heterogeneity could be removed by creating WES- and WGS-specific signatures. This should ideally rely on the most comprehensive inventories obtained by WGS, because this also maximizes the ability to obtain insight into the underlying biological mechanisms. For clinical use, however, refitting of predefined signatures on WES data is likely feasible and more cost-efficient, which would make mutational signature analysis more broadly applicable. Second, a number of the current signatures are identified in only a few genomes at low contributions [8]. Their relevance should be substantiated before they are used in refitting approaches, because such signatures may mask the contributions of other signatures due to overlapping features. Likewise, signatures observed in single cohorts likely represent artifacts due to sequencing errors or from inadequate somatic mutation calling pipelines [99]. Consequently, the identification of such artifactual signatures makes it valid to question and optimize the sensitivity and specificity of the mutation calling strategy. Alternatively, artifactual signatures can be included to capture predefined false positive mutations as for example in single cell sequencing that generates numerous T > C mutations [100]. Third, not all mutational signatures will lead to targeting approaches or clinical advice. It is arguably unlikely that age-related CS-1, which is present in approximately 70% of all cancer types, can be translated to any form of prevention or treatment. Fourth, the accuracy of mutational signature extraction decreases when a multitude of mutational processes are or have been active in a sample, when low numbers of mutations are present (e.g. pediatric cancers and adult acute myeloid leukemia (AML) [101]), and when mutational signatures are relatively similar. Fifth, it is preferable to distinguish historical mutational processes from those that are presently ongoing to identify a signature-based treatment. For example, targeting APOBEC overactivity, a process that is known to operate transiently, solely on the presence of its signature will not necessarily affect patient survival. Likewise, subclones within tumors that have lost the activity of certain mutational processes will still contain their characteristic signatures within the genome. Moreover, subclones that have become the dominant clone during cancer recurrence after the first stages of treatment will still show their historic mutational signatures. In a diagnostic setting, mutational processes may be classified as historical

or ongoing by analyzing samples from serial biopsies or biopsies from different sites within the tumor. Alternatively, active signatures can be characterized computationally by focusing on subclonal variants, because they are considered to originate from recent processes in local portions of the cancer. More sophisticated computational strategies also exist to assess the evolutionary history of mutational processes [102, 103].

It is important to mention that signatures of mutational classes other than base substitutions have enjoyed less attention. This is partly due to the known lower sensitivity and specificity of current algorithms used to call indel and structural variant mutations, which results in noisier data and more challenging extraction of biologically relevant signatures, as well as the higher complexity of defining other signatures [9]. The context of these signatures includes features beyond neighboring nucleotides, such as length, location, repeat engagement, copy-number changes, involvement of microhomology, and other biologically relevant attributes. Regarding indels, two distinct informative signatures were defined by Stratton and colleagues in breast cancer [9, 15]. The first indel signature is characterized by small indels (1–5 bp) flanked by short tandem repeats (STRs), and the second is characterized by larger indels (up to 50 bp) present in short stretches of identical sequences at the breakpoints (microhomology). Regarding structural rearrangements, six signatures (RS1-RS6) based on rearrangement type (duplications, deletions, translocations, inversions), degree of clustering, and size have been identified by analyzing 560 breast cancer genomes [22]. These indel and RS signatures are also found in liver cancer [94], highlighting the robustness of these preliminary indel and RS signatures. More signatures are likely to be recognized in the near future as techniques to identify indels and rearrangements develop and as cancer genomes are more systematically analyzed [104]. Additional relevant parameters may be incorporated into signatures in the future as well, including genomic features such as transcriptional strand bias [8, 105], replication timing [106, 107], genomic position [77], chromatin organization [108, 109], and other relevant genomic features [110]. For instance, heterogeneity in mutation rate has been observed within a single gene that is associated with higher mismatch repair activity in exonic regions [111], and clustered mutation signatures are related to variable APOBEC activity and tobacco smoking [77]. The inclusion of such parameters into patterns increases the resolution of mutational signatures to distinguish different processes. These parameters could also be clinically meaningful as stand-alone signatures, such as indels in STRs to identify MMR deficiency [112]. Which parameters must be incorporated into signatures and which can stand alone is a question that could potentially be addressed by feature correlation analyses of very

large cancer genomics datasets. Furthermore, incorporating biases into signatures enhances the power of mutational signature analysis to detect underlying mutational processes.

The algorithmic approach behind mutational signature analysis still requires further development. A recent study validated a number of peer-reviewed mutational signature frameworks and found large variation in signature exposures, of which NMF gave on average the largest decomposition error [113]. Furthermore, NMF relies upon large cohorts of cancer genomes to accurately extract signatures and cannot efficiently analyze samples with high mutational load. Hence, a growing number of bioinformatics studies are attempting to address the shortcomings of NMF [16] by proposing and testing different mathematical approaches for problems such as defining the optimal number of signatures in a sample [114–120]. Alternatively, after a complete set of mutational signatures has been verified, the contribution of these predefined signatures (e.g., those currently recorded in the COSMIC database) could be refitted on the genomic data of a single patient [117, 119]. The latter strategy might prove faster and more cost-effective [43] and, most importantly, is applicable at the single patient level, which is a requirement for use in a clinical diagnostic setup. Methods for refitting known signatures to mutation inventories are still in their infancy, and are faced with challenges due to the overlapping characteristics of signatures, making it difficult to assign individual mutations to specific signatures. Hence, additional specific genomic features (e.g. broader mutation context, strand biases, association with functional elements) exclusively linked to a signature might be crucial to accurately asses the contribution of highly similar signatures and could simultaneously make refitting approaches more accurate.

Also, not all forces driving tumorigenesis might be detectable by DNA mutation analyses. Epigenetic modifications are another important cancer driver mechanism, but such alterations are not detected by routine WGS. It has been suggested that epigenetic changes, as detected by other targeted or genome-wide techniques, could be integrated into mutational signature analysis if need be [11]; however, no framework has been published yet.

Feasibility and costs

Despite the recent advances in DNA sequencing technology and the consequent wave of studies using mutational patterns, diagnostic application of mutational signatures is still at an early stage of development. Certain mutational signatures can be linked to mutational processes and, via this route, to a treatment plan. To date, however, studies on how mutational signature-based subtyping translates to treatment response are largely absent. Studies using

HRDetect or stratifying studies on the basis of mutational signatures do demonstrate a correlation with therapy response [11, 43, 44], but these studies were performed retrospectively. Therefore, the major challenge for mutational signature analysis will be to predict treatment response in a prospective study.

In addition, mutational signature analysis requires NGS data to accurately identify somatic mutations, preferably from WGS data with sufficient sequencing depth, accompanied by a matched healthy sample. WGS-derived data contain 20–50 times more mutations than do data from whole exomes [121, 122]. Hence, the decomposition of a patient's mutational profile into de novo signatures using WES data may generate unstable signatures, as discussed above. However, Polak et al. [12] successfully detected BRCAness in the TCGA WES-derived dataset using an optimal threshold of 37 CS-3 associated mutations (AUC = 0.82). Therefore, refitting on robust mutational signatures and optimizing threshold levels may well work with only exome sequencing data of the diagnostic sample [98, 100]. Regarding sequencing depth, only a small drop in sensitivity was observed in the WGS breast cancer analysis when data with a 30-fold read depth was down-sampled to a 10-fold read depth (r = 0.96), with a remaining sensitivity of 86% for low-coverage sequencing data [12]. Similarly, a simulated 10-fold read depth could be successfully used to identify the dominant signature for EAC-patient stratification [43], although the required read depths will also strongly depend on the percentage of tumor cells in the sample and the tumor heterogeneity. Furthermore, the current somatic calling pipeline demands a matched healthy DNA sequence to distinguish somatic mutations from germline variants. However, the establishment of comprehensive population resources and well-trained computer models could potentially overcome this requirement without losing the detection power of mutational signature analysis.

Studies presenting the feasibility of mutational signature analysis for cancer patients have mostly used high-quality DNA extracted from fresh-frozen biopts. However, in clinical practices, such specimens are routinely fixed in formalin and paraffin-embedded (FFPE) for histopathological diagnosis, which lowers DNA quality [123]. Nevertheless, HRDetect (using 30-fold read depth) sustained high probability using FFPE tissues, indicating that mutational signature analysis may work in the current framework of molecular pathology. However, low-exposure signatures such as CS-3 and CS-5 might be lost in FFPE-induced noise [11].

Overall, it is difficult to draw conclusions on the cost-effectiveness of mutational signature analysis at this moment, although it is clear that costs for WGS are still clearly prohibitive for routine application in most clinical studies. However, when the potential of WGS to replace

the multifactorial testing of mutated genes and to allow better patient stratification is met, cost-effectiveness could likely be reached, because WGS costs are only a fraction of total clinical study costs or the costs of current novel targeted treatments. In addition, with the decreasing costs of NGS, full commitment to WGS might be less of an issue at some point in the future.

Clinical trials

Mutational signature analyses have been applied in research, but not yet in the clinical setting. Currently, our theoretical understanding of the mechanisms by which mutational signatures accumulate is still relatively rudimentary. However, the findings that are gained by WGS analysis open up the question of when WGS analyses will enter routine clinical cancer care. Therefore, examination of the mutational landscapes in clinical trials exploring the accuracy of this approach in a wide range of cancer types are a pertinent next objective. So far, only a very limited number of clinical trials (as reported on clinicaltrials.gov) have been initiated to examine the clinical relevance of mutational signatures.

The potential therapeutic efficacy of the PARP inhibitor olaparib in *BRCA*-mutated tumors has been assessed in clinical trials in breast cancer (NCT00494234 – completed), ovarian cancer (NCT00494442 – completed; NCT00753545 – completed; NCT00679783 – completed), prostate and pancreatic cancer (NCT01078662 – completed, NCT02677038 – recruiting; NCT02184195 – recruiting) and has been approved by the FDA in 2014 [124]. Currently, other PARP inhibitors are being tested for BRCA-deficient cancers in clinical settings, such as veliparib (NCT01149083) and rucaparib (NCT02855944), and platinum-based chemotherapy has been tested in prostate cancer (NCT01289067). However, these patients were mostly screened using targeted assays for germline and somatic BRCA mutations. The development of a companion diagnostic biomarker that relies on signatures (such as HRDetect) could guide treatment of HRR-deficient cancer types beyond those carrying *BRCA* mutations in the cancer types discussed above, and thus increase the target population. In this context, one trial (NCT01042379) investigated a BRCA-signature from gene expression data that was developed within the EU FP7 RATHER project, which showed promise in predicting the response to PARP inhibitor veliparib in combination with carboplatin [125]. However, the prognostic and diagnostic value of BRCA-associated signatures from somatic mutations remains to be assessed through a prospective clinical trial, with participants being selected based on the mutational signatures of their tumor.

One trial (NCT02710396) is currently recruiting patients to explore the mutational smoking signature as a potential biomarker in advanced non-small cell lung cancer treated with pembrolizumab. This PD1-blocking agent was FDA-approved in May 2017 for cancer patients diagnosed with microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR) cancers. Currently, MSI detection depends on a small number of known microsatellite loci or mismatch repair genes, and has limited reliability [126]. However, NGS data can offer highly accurate detection of MSI [127, 128]. Pembrolizumab was the first FDA-approved cancer treatment solely based on a genetic biomarker, rather than in combination with a primary tumor type. This decision opens up the route for additional biomarkers that focus on genomic profiles. In this context, a clinical trial (NCT02750657) has been set up to study the potential of mutational signature analysis for better treatment selection in PDAC, which is currently recruiting patients and might prove important for realizing the diagnostic potential of mutational signature analysis [44].

Conclusion

In conclusion, cancer diagnosis may benefit from the implementation of mutational signature analysis, which is complementary to existing diagnostic approaches such as analyses of driver mutations in oncogenes and tumor suppressors. The identification of HRR deficiency in breast cancer and other cancers suggest the potential for a broader application of mutational signature analysis in different cancer types. Moreover, the detection of additional signatures suggests that similar developments could occur in the diagnosis of a broader range of DNA repair defects. Mutational signatures are proving to be clinically useful biomarkers for a growing range of cancer types, and signatures have already been shown to be useful for prognosis in several studies, such as the prediction of responses to conventional chemotherapy, targeted therapy, and immunotherapy approaches. Moreover, mutational signatures are found to be powerful biomarkers for the identification of hereditary cancer syndromes, providing opportunities for cancer prevention, monitoring, and early detection strategies.

Despite these promising results, mutational signature analysis will need further research to define universal reference signatures based on all types of mutational events and relevant genomic features, as well as to delineate the underlying mutational processes. This will require analyses of extensive, and more diverse, cancer genome sequencing datasets, as well as the targeted manipulation or perturbation of experimental models. Moreover, it is important that prospective clinical trials are undertaken to assess the effectiveness and accuracy of mutational signature analyses in predicting response to therapy. Finally, for patients to benefit from these developments, transparency regarding technical advances in algorithms and sharing of methods and data are

Hoeck *et al. BMC Cancer* (2019) 19:457

Page 11 of 14

imperative for the timely and responsible transfer of mutational signature analyses from the research domain to the clinical setting.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Center for Molecular Medicine and Oncode Institute, University Medical Centre Utrecht, Heidelberglaan 100, 3584CX Utrecht, The Netherlands. [2]Princess Máxima Center for Pediatric Oncology and Oncode Institute, Heidelberglaan 25, 3584CS Utrecht, The Netherlands. [3]Hartwig Medical Foundation, Science Park 408, 1098XH Amsterdam, The Netherlands.

References
1. Hudson TJ, Anderson W, Aretz A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. Nature. 2010;464:993–8.
2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458: 719–24.
3. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013;153: 17–37.
4. Bernards R. It's diagnostics, stupid. Cell. 2010;141:13–7.
5. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23:703–13.
6. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371–6.
7. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321–7.
8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S a JR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–21.
9. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014;15:585–98.
10. Vanderstichele A, Busschaert P, Olbrecht S, Lambrechts D, Vergote I. Genomic signatures as predictive biomarkers of homologous recombination deficiency in ovarian cancer. Eur J Cancer. 2017;86:5–14.
11. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23:517–25.
12. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017;49:1476–86.
13. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47:1402–7.
14. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354:618–22.
15. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149:979–93.
16. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human Cancer. Cell Rep. 2013;3:246–59.
17. Nik-Zainal S, Morganella S. Mutational signatures in breast Cancer: the problem at the DNA level. Clin Cancer Res. 2017;23:2617–29.
18. Giglia-Mari G, Sarasin A. TP53 mutations in human skin cancers. Hum Mutat. 2003;21:217–28.
19. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21:7435–51.
20. Hollstein M, Moriya M, Grollman AP, Olivier M. Analysis of TP53 mutation spectra reveals the fingerprint of the potent environmental carcinogen, aristolochic acid. Mutat Res. 2013;753:41–9.
21. Hollstein M, Sidransky D, Vogelstein B, Harris C. p53 mutations in human cancers. Science. 1991;253(5015):49–53.
22. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534:1–20.
23. Huang KK, Jang KW, Kim S, Kim HS, Kim S-M, Kwon HJ, et al. Exome sequencing reveals recurrent REV3L mutations in cisplatin-resistant squamous cell carcinoma of head and neck. Sci Rep. 2016;6:19552.
24. Bueno R, Stawiski EW, Goldstein LD, Durinck S, De Rienzo A, Modrusan Z, et al. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. Nat Genet. 2016;48:407–16.
25. Li X, Wu WKK, Xing R, Hwong S, Liu Y, Fang X, et al. Distinct subtypes of gastric cancer defined by molecular characterization include novel mutational signatures with prognostic capability. Cancer Res. 2016;76:1724–32.
26. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, et al. A specific mutational signature associated with DNA 8-Oxoguanine persistence in MUTYH-defective colorectal Cancer. EBioMed. 2017;20:39–49.
27. Campbell PJ, Getz G, Stuart JM, Korbel JO, Stein LD. Pan-cancer analysis of whole genomes. bioRxiv. 162784. https://doi.org/10.1101/162784.

28. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455:1061–8.

29. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010;463:191–6.

30. Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science. 2017;238:eaao3130.

31. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361:557–68.

32. Hu X, Huang W, Fan M. Emerging therapies for triple-negative breast cancer. J Hematol Oncol. 2017;10:1–17.

33. Kanjanapan Y, Lheureux S, Oza AM. Niraparib for the treatment of ovarian cancer. Expert Opin Pharmacother. 2017;18:631–40.

34. Melinda LT, Kirsten MT, Julia R, Bryan H, Gordon BM, Kristin CJ, et al. Homologous recombination deficiency (hrd) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. Clin Cancer Res. 2016;22:3764–73.

35. Pennington KP, Walsh T, Harrell MI, Lee MK, Pennil CC, Rendi MH, et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. Clin Cancer Res. 2014;20:764–75.

36. Gerratana L, Fanotto V, Pelizzari G, Agostinetto E, Puglisi F. Do platinum salts fit all triple negative breast cancers? Cancer Treat Rev. 2016;48:34–41.

37. Watkins JA, Irshad S, Grigoriadis A, Tutt AN. Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. Breast Cancer Res. 2014;16:1–11.

38. Graeser M, McCarthy A, Lord CJ, Savage K, Hills M, Salter J, et al. A marker of homologous recombination predicts pathologic complete response to neoadjuvant chemotherapy in primary breast cancer. Clin Cancer Res. 2010; 16:6159–68.

39. Sachs N, de Ligt J, Kopper O, Gogola E, Bounova G, Weeber F, et al. A living biobank of breast Cancer organoids captures disease heterogeneity. Cell. 2018;172:373–86.

40. Metcalfe KA, Lynch HT, Ghadirian P, Tung N, Olivotto IA, Foulkes WD, et al. The risk of ovarian cancer after breast cancer in BRCA1 and BRCA2 carriers. Gynecol Oncol. 2005;96:222–6.

41. Goggins M, Schutte M, Lu J, CA M, Weinstein CL, Petersen GM, et al. Germline BRCA2 gene mutations in patients with apparently sporadic pancreatic carcinomas. Cancer Res. 1996;56:5360–4.

42. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer. 2016;16:110–20.

43. Secrier M, Li X, de Silva N, Eldridge MD, Contino G, Bornschein J, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. Nat Genet. 2016;48(10):1131–41.

44. Connor AA, Denroche RE, Jang GH, et al. Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma. JAMA Oncol. 2017;3(6):774–83.

45. Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531:47–52.

46. Harfe BD, Jinks-Robertson S. DNA mismatch repair and genetic instability. Annu Rev Med. 2000;34:359–99.

47. Cortes-Ciriano I, Lee S, Park W-Y, Kim T-M, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun. 2017;8:1–12.

48. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015; 372:2509–20.

49. Bouffet E, Larouche V, Campbell BB, Merico D, de Borja R, Aronson M, et al. Immune checkpoint inhibition for Hypermutant glioblastoma Multiforme resulting from germline Biallelic mismatch repair deficiency. J Clin Oncol. 2016;34:2206–11.

50. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. Nat Med. 2016;22:1342–50.

51. Humphris JL, Patch AM, Nones K, Bailey PJ, Johns AL, McKay S, et al. Hypermutation In Pancreatic Cancer. Gastroenterology. 2017;152:68–74 e2.

52. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch-repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;6733:1–11.

53. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;357:409–13.

54. An Q, Robins P, Lindahl T, Barnes DE. C --> T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. EMBO J. 2005;24:2205–13.

55. Smart DJ, Chipman JK, Hodges NJ. Activity of OGG1 variants in the repair of pro-oxidant-induced 8-oxo-2′-deoxyguanosine. DNA Repair (Amst). 2006;5: 1337–45.

56. Alsøe L, Sarno A, Carracedo S, Domanska D, Dingler F, Lirussi L, et al. Uracil accumulation and mutagenesis dominated by cytosine deamination in CpG dinucleotides in mice lacking UNG and SMUG1. Sci Rep. 2017;7:1–14.

57. Weren RDA, Ligtenberg MJL, Kets CM, de Voer RM, Verwiel ETP, Spruijt L, et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet. 2015;47:668–71.

58. Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol. 2017;242:10–5.

59. Ohno M, Sakumi K, Fukumura R, Furuichi M, Iwasaki Y, Hokama M, et al. 8-Oxoguanine causes spontaneous De novo germline mutations in mice. Sci Rep. 2014;4:1–9.

60. Guo J, Hanawalt PC, Spivak G. Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells. Nucleic Acids Res. 2013;41:7700–12.

61. Hanawalt PC, Spivak G. Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol. 2008;9:958–70.

62. Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Tiao G, et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat Genet. 2016;48:600–6.

63. Olaussen KA, Dunant A, Fouret P, Brambilla E, André F, Haddad V, et al. DNA repair by ERCC1 in non–small-cell lung Cancer and cisplatin-based adjuvant chemotherapy. N Engl J Med. 2006;355:983–91.

64. Van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, et al. Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. Cancer Discov. 2014;4:1140–53.

65. Stubbert LJ, Smith JM, McKay BC. Decreased transcription-coupled nucleotide excision repair capacity is associated with increased p53- and MLH1-independent apoptosis in response to cisplatin. BMC Cancer. 2010;10:207.

66. Jager M, Blokzijl F, Kuijk E, Bertl J, Vougioukalaki M, Janssen R, Besselink N, Boymans S, de Ligt J, Pedersen JS, Hoeijmakers J, Pothof J, van Boxtel R, Cuppen E. Deficiency of nucleotide excision repair explains mutational signature observed in cancer. bioRxiv. 2018:221168. https://www.biorxiv.org/content/10.1101/221168v2.

67. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer. 2014;14:786–800.

68. Fox EJ, Loeb LA. Lethal mutagenesis: targeting the Mutator phenotype in Cancer. Semin Cancer Biol. 2010;20:353–9.

69. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC enzymes: mutagenic fuel for Cancer evolution and heterogeneity. Cancer Discov. 2015;5:704–12.

70. Sieuwerts AM, Willis S, Burns MB, Look MP, Van GMEM, Schlicker A, et al. Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. Horm Cancer. 2014;5:405–13.

71. Law EK, Sieuwerts AM, LaPara K, Leonard B, Starrett GJ, Molan AM, et al. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. Sci Adv. 2016;2:e1601737.

72. Nakamura H, Arai Y, Totoki Y, Shirota T, Elzawahry A, Kato M, et al. Genomic spectra of biliary tract cancer. Nat Genet. 2015;47:1003–10.

73. Cifola I, Lionetti M, Pinatel E, Todoerti K, Mangano E, Pietrelli A, et al. Whole-exome sequencing of primary plasma cell leukemia discloses heterogeneous mutational patterns. Oncotarget. 2015;6:17543–58.

74. Yu W, McPherson JR, Stevenson M, Van Eijk R, Heng HL, Newey P, et al. Whole-exome sequencing studies of parathyroid carcinomas reveal novel PRUNE2 mutations, distinctive mutational spectra related to APOBEC-catalyzed DNA mutagenesis and mutational enrichment in kinases associated with cell migration and invasion. J Clin Endocrinol Metab. 2015; 100:E360–4.

75. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat Genet. 2014;46:487–91.

Hoeck *et al. BMC Cancer*   (2019) 19:457

Page 13 of 14

76. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47:1067–72.

77. Supek F, Lehner B, Supek F, Lehner B, Supek F, Lehner B, et al. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. Cell. 2017;170:534–547.e23.

78. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC CancerBase no. 11. International Agency for Research on Cancer. 2013. http://globocan.iarc.fr. Accessed 26 Jun 2017.

79. Fackenthal JD, Olopade OI. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nat Rev Cancer. 2007;7:937–48.

80. Su SC, Lin CW, Liu YF, Fan WL, Chen MK, Yu CP, et al. Exome sequencing of Oral squamous cell carcinoma reveals molecular subgroups and novel therapeutic opportunities. Theranostics. 2017;7:1088–99.

81. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017;541:359–64.

82. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45:1127–33.

83. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nat Genet. 2013;45:478–86.

84. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blocade in non-small cell lung cancer. Science. 2015;348:124–9.

85. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371:2189–99.

86. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Immune Checkpoint Blockade. Science. 2016;351:1463–9.

87. Romero-Laorden N, Castro E. Inherited mutations in DNA repair genes and cancer risk. Curr Probl Cancer. 2017;41:251–64.

88. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. Nat Rev Genet. 2016;17:93–108.

89. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive analysis of Hypermutation in human Cancer. Cell. 2017;171: 1042–56 e10.

90. Eng C, Hampel H, de la Chapelle A. Genetic testing for cancer predisposition. Annu Rev Med. 2000;52:371–400.

91. Sampson JR, Jones S, Dolwani S, Cheadle JP. MutYH (MYH) and colorectal cancer. Biochem Soc Trans. 2005;33(Pt 4):679–83.

92. Tan MH, Mester J, Peterson C, Yang Y, Chen JL, Rybicki LA, et al. A clinical scoring system for selection of patients for pten mutation testing is proposed on the basis of a prospective study of 3042 probands. Am J Hum Genet. 2011;88:42–56.

93. Pavlidis N, Fizazi K. Carcinoma of unknown primary (CUP). Crit Rev Oncol Hematol. 2009;69:271–8.

94. Letouzé E, Shinde J, Renault V, Couchy G, Blanc JF, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun. 2017;8:1315.

95. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature. 2015;521:489–94.

96. Dietlein F, Eschner W. Inferring primary tumor sites from mutation spectra: a meta-analysis of histology-specific aberrations in cancer-derived cell lines. Hum Mol Genet. 2014;23:1527–37.

97. Marquard AM, Birkbak NJ, Thomas CE, Favero F, Krzystanek M, Lefebvre C, et al. TumorTracer: a method to identify the tissue of origin from the somatic mutations of a tumor specimen. BMC Med Genet. 2015;8:58.

98. Jiao W, Atwal G, Polak P, Karlic R, Cuppen E, Danyi A, Ridder J, van Herpen C, Lolkema MP, Steeghs N, Getz G, Morris QD, Stein LD. PCAWG Pathology & Clinical Correlates Working Grp, ICGC/TCGA Pan-cancer Analysis of Whole Genomes Net A deep learning system can accurately classify primary and metastatic cancers based on patterns of passenger mutations. bioRxiv. 214494.

99. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. Nat Commun. 2015;6:10001.

100. Petljak M, Alexandrov LB, Brammeld JS, et al. Characterizing mutational signatures in human Cancer cell lines reveals episodic APOBEC mutagenesis. Cell. 2019;176:1282–94.

101. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150:264–78.

102. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci. 2010;107: 16910–5.

103. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149:994–1007.

104. Li Y, Roberts ND, Weischenfeldt J, Wala JA, Shapira O, Schumacher SE, Khurana E, Korbel J, Imielinski M, Beroukhim R, Campbell PJ, on behalf of the PCAWG-Structural Variation Working Group, and the PCAWG Network. Patterns of structural variation in human cancer. bioRxiv. 181339.

105. Popadin K, Seplyarskiy VB, Soldatov RA, Popadin KY, Antonarakis SE, Bazykin GA, et al. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. Genome Res. 2016; 26(2):174–82.

106. Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. Am J Hum Genet. 2012;91:1033–40.

107. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nat Genet. 2009;41:393–5.

108. Polak P, Karlic R, Koren A, Thurman R, Sandstrom R, Lawrence MS, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature. 2015;518:360–4.

109. Schuster-Böckler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature. 2012;488:504–7.

110. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. Nat Rev Genet. 2011;12:756–66.

111. Frigola J, Sabarinathan R, Mularoni L, Muinõs F, Gonzalez-Perez A, López-Bigas N. Reduced mutation rate in exons due to differential mismatch repair. Nat Genet. 2017;49:1684–92.

112. Lin EI, Tseng L-H, Gocke CD, Reil S, Le DT, Azad NS, et al. Mutational profiling of colorectal cancers with microsatellite instability. Oncotarget. 2015;6:42334–44.

113. Huang X, Wojtowicz D, Przytycka TM. Detecting presence of mutational signatures in cancer with confidence. Bioinformatics. 2018;34:330–7.

114. Rosales RA, Drummond RD, Valieris R, Dias-Neto E, Da Silva IT. signeR: An empirical Bayesian approach to mutational signature discovery. Bioinformatics. 2017;33:8–16.

115. Kakushadze Z, Yu W. Factor models for cancer signatures. Physica A. 2016; 462:527–59.

116. Shiraishi Y, Tremmel G, Miyano S, Stephens M. A simple model-based approach to inferring and visualizing Cancer mutation signatures. PLoS Genet. 2015;11:1–21.

117. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:1–11.

118. Gehring JS, Fischer B, Lawrence M, Huber W. Somatic Signatures: Inferring mutational signatures from single-nucleotide variants. Bioinformatics. 2015; 31:3673–5.

119. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. Mutational Patterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10:33.

120. Baez-Ortega A, Gori K. Computational approaches for discovery of mutational signatures in cancer. Brief Bioinform. 2019;20(1):77–88.

121. Loeb LA. Human cancers express mutator phenotypes: origin, consequences and targeting. Nat Rev Cancer. 2011;11:450–7.

122. Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. Nat Commun. 2015;6:8683.

123. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol. 2002;161:1961–71.

124. Kim G, Ison G, McKee AE, Zhang H, Tang S, Gwise T, et al. FDA approval summary: Olaparib monotherapy in patients with deleterious germline BRCA-mutated advanced ovarian Cancer treated with three or more lines of chemotherapy. Clin Cancer Res. 2015;21:4257–61.

125. Severson TM, Wolf DM, Yau C, Peeters J, Wehkam D, Schouten PC, et al. The BRCA1ness signature is associated significantly with response to PARP inhibitor treatment versus control in the I-SPY 2 randomized neoadjuvant setting. Breast Cancer Res. 2017;19:1–9.

126. Umar A, Boland CR, Terdiman JP, Syngal S, Chapelle AD, Ruschoff J, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal Cancer (Lynch syndrome) and microsatellite instability. JNCI. 2004;96:261–8.
127. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. Clin Chem. 2014;60:1192–9.
128. Lu Y, Soong TD, Elemento O. A novel approach for characterizing microsatellite instability in Cancer cells. PLoS One. 2013;8:e63056.