

ARTICLE OPEN



Multicentric pilot study to standardize clinical whole exome sequencing (WES) for cancer patients

Michael Menzel ^{1,2}, Stephan Ossowski^{3,4,5,23}, Sebastian Kral ^{6,7,23}, Patrick Metzger ^{7,8,23}, Peter Horak ^{2,9,10,23}, Ralf Marienfeld^{11,12,23}, Melanie Boerries ^{7,8,13,14}, Steffen Wolter ^{6,7}, Markus Ball ^{1,2}, Olaf Neumann ^{1,2}, Sorin Armeanu-Ebinger^{3,4}, Christopher Schroeder ^{3,4}, Uta Matysiak^{6,7}, Hannah Goldschmid^{1,2}, Vincent Schipperges^{7,8}, Axel Fürstberger ^{11,12,15}, Michael Allgäuer^{1,2}, Timo Eberhardt^{11,12}, Jakob Niewöhner¹¹, Andreas Blaumeiser^{7,8,14}, Carolin Ploeger^{1,2}, Tobias Bernd Haack^{3,4}, Timothy Kwang Yong Tay^{1,2,16}, Olga Kelemen^{3,4}, Thomas Pauli^{7,8}, Martina Kirchner^{1,2}, Klaus Kluck ^{1,2}, Alexander Ott^{3,4}, Marcus Renner^{1,2,9}, Jakob Admard ^{3,4}, Axel Gschwind ^{3,4}, Silke Lassmann^{6,7}, Hans Kestler^{11,12}, Falko Fend ¹⁷, Anna Lena Illert^{18,19,20}, Martin Werner^{6,7,14}, Peter Möller¹¹, Thomas Theodor Werner Seufferlein^{12,21}, Nisar Malek^{4,22}, Peter Schirmacher^{1,2,10}, Stefan Fröhling ^{1,2,9,10,24}, Daniel Kazdal ^{1,2,24}, Jan Budczies ^{1,2,10,24} and Albrecht Stenzinger ^{1,2,10,24}

A growing number of druggable targets and national initiatives for precision oncology necessitate broad genomic profiling for many cancer patients. Whole exome sequencing (WES) offers unbiased analysis of the entire coding sequence, segmentation-based detection of copy number alterations (CNAs), and accurate determination of complex biomarkers including tumor mutational burden (TMB), homologous recombination repair deficiency (HRD), and microsatellite instability (MSI). To assess the inter-institution variability of clinical WES, we performed a comparative pilot study between German Centers of Personalized Medicine (ZPMs) from five participating institutions. Tumor and matched normal DNA from 30 patients were analyzed using custom sequencing protocols and bioinformatic pipelines. Calling of somatic variants was highly concordant with a positive percentage agreement (PPA) between 91 and 95% and a positive predictive value (PPV) between 82 and 95% compared with a three-institution consensus and full agreement for 16 of 17 druggable targets. Explanations for deviations included low VAF or coverage, differing annotations, and different filter protocols. CNAs showed overall agreement in 76% for the genomic sequence with high wet-lab variability. Complex biomarkers correlated strongly between institutions (HRD: 0.79–1, TMB: 0.97–0.99) and all institutions agreed on microsatellite instability. This study will contribute to the development of quality control frameworks for comprehensive genomic profiling and sheds light onto parameters that require stringent standardization.

npj Precision Oncology (2023)7:106; <https://doi.org/10.1038/s41698-023-00457-x>

INTRODUCTION

Scientific advances and technological developments over the past few decades led to a growing number of drugs that are available for treatment of cancer patients¹. The paradigm of precision oncology, which is based on a mechanistic understanding of tumor biology and fine granular diagnostic profiling, is supported by the majority of clinical trials². Furthermore, the clinical relevance of complex biomarkers, like microsatellite instability (MSI), tumor mutational burden (TMB), and homologous recombination deficiency (HRD), is increasing since they have been shown to be meaningful predictive biomarkers for patient stratification. As a result, there is movement towards more comprehensive DNA

sequencing in routine clinical care, evident in the shift from the use of smaller (few to 100 genes) to larger (around 1 Mb) targeted panels in Next Generation Sequencing (NGS)^{3–6}. While this diagnostic approach is generally supported by reimbursement schemes in national health care systems, whole exome sequencing (WES) and whole genome sequencing (WGS) are currently employed mainly in dedicated research programs aimed at comprehensive screening and identification of novel biomarkers and druggable targets^{7–9}. But, with increasing sequencing capabilities and decreasing sequencing costs, WES and possibly even WGS have the potential to enter the clinical setting^{10–15}. Compared to panel sequencing, WES offers the opportunity to (i)

¹Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany. ²Center for Personalized Medicine (ZPM), Heidelberg, Germany. ³Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany. ⁴Center for Personalized Medicine (ZPM), Tübingen, Germany. ⁵Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany. ⁶Institute for Surgical Pathology, Medical Center, University of Freiburg, Freiburg, Germany. ⁷Center for Personalized Medicine (ZPM), Freiburg, Germany. ⁸Institute of Medical Bioinformatics and Systems Medicine (IBSM), Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ⁹Division of Translational Medical Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. ¹⁰German Cancer Consortium (DKTK), Heidelberg, Germany. ¹¹Institute of Pathology, University Hospital Ulm, Ulm, Germany. ¹²Center for Personalized Medicine (ZPM), Ulm, Germany. ¹³Comprehensive Cancer Center Freiburg (CCCF), Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. ¹⁴German Cancer Consortium (DKTK) Partner Site Freiburg, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁵Institute of Medical Systems Biology, Ulm University, Ulm, Germany. ¹⁶Department of Anatomical Pathology, Singapore General Hospital, Singapore, Singapore. ¹⁷Institute of Pathology and Neuropathology, University Hospital Tübingen, Tübingen, Germany. ¹⁸Department of Medicine I, Medical Center-University of Freiburg, Faculty of Medicine, University of Freiburg, 79085 Freiburg, Germany. ¹⁹Medical Department for Hematology and Oncology, Klinikum Rechts der Isar, Technische Universität München, 80333 Munich, Germany. ²⁰German Cancer Consortium (DKTK) Partner Site Munich, and German Cancer Research Center (DKFZ), Heidelberg, Germany. ²¹Department of Internal Medicine I, University of Ulm, Ulm, Germany. ²²Department of Internal Medicine I, University Hospital Tübingen, Tübingen, Germany. ²³These authors contributed equally: Stephan Ossowski, Sebastian Kral, Patrick Metzger, Peter Horak, Ralf Marienfeld. ²⁴These authors jointly supervised this work: Stefan Fröhling, Daniel Kazdal, Jan Budczies, Albrecht Stenzinger. [✉]email: jan.budczies@med.uni-heidelberg.de; albrecht.stenzinger@med.uni-heidelberg.de

comprehensively cover the coding sequence so that either the entire exome or virtual gene panels can be investigated in the light of present and future diagnostic needs, (ii) facilitate comparability and standardization across different laboratories by avoiding bias of custom gene panels which may differ in design and require repetitive updates on gene content, and (iii) accurately measure (instead of estimate) complex biomarkers such as TMB, HRD, and MSI^{16–19}.

WES focusing on cancer diagnostics has been investigated in various single center studies^{9,12,20–22}, as well as in multi-center studies focusing on singular metrics including variant detection^{23–28}, panel-based somatic variant detection^{29–33}, CNAs^{34–37}, or complex biomarkers^{38–41}. However, WES studies that assess the use of real-world clinical tissue samples and investigate the reproducibility of diagnostic results across different diagnostic institutes covering the entire workflow from laboratory analysis to bioinformatic evaluation are scarce. To fill this gap, the Centers of Personalized Medicine of Baden-Wuerttemberg (ZPM)⁴² embarked on a joint pilot study to pave the way for the use of WES in routine cancer diagnostics.

Tumor and matched normal DNA samples from 30 patients covering a variety of cancer entities and molecular alteration types were distributed to four participating centers, located in Heidelberg, Freiburg, Ulm and Tübingen. The samples were sequenced in the respective molecular laboratories according to their protocols and analyzed using the local bioinformatic pipelines of five different institutions at the four ZPM partner sites. The institutions reported a set of pre-defined molecular biomarkers: somatic mutations and CNAs in a list of relevant genes as well as the complex biomarkers HRD, TMB, and MSI. Inter-institution comparisons showed a high concordance for the somatic variant calls, CNA calls and the three complex biomarkers. An in-depth analysis was performed to reveal the causes for differences between the results of the participating institutions.

RESULTS

Paired tumor and normal DNA samples from 30 patients were subjected to WES at four ZPM laboratories with subsequent data analysis performed at five ZPM bioinformatic institutions (Fig. 1).

DNA was originally derived from fresh-frozen samples that had been analyzed by WES within the multicenter MASTER (Molecularly Aided Stratification for Tumor Eradication)⁹ program of the German Cancer Research Center (DKFZ), the National Center for Tumor Diseases (NCT) Heidelberg/Dresden, and the German Cancer Consortium (DKTK) before. The cohort consisted of a variety of rare cancer types including carcinomas, sarcomas, and other tumor entities selected to represent a wide variety of molecular alterations (Supplementary Table 1). Sequencing and bioinformatic analysis were performed agnostic of the tumor entity.

All DNA samples were successfully sequenced by all participating centers utilizing their local enrichment-kits, sequencers, and wet-lab protocols (Supplementary Table 2). Quality measures calculated for each of the samples resulted in on-target rates (percentage of bases in the interrogated region) of 41–83% and mean coverages after deduplication of 56–504x (Supplementary Figure 1). Concordance analysis between the five institutions of the four centers was performed in a predefined set of 494 oncological relevant genes (Supplementary Table 3).

Somatic variant calling

Somatic variants were compared between institutions based on their chromosomal position and alteration at the DNA level. Only exact matches were accepted as the same variant. In the predefined gene set, a total of 960 somatic variant calls including 804 single-nucleotide variants (SNV) and 156 deletions/insertions were made by the five participating institutions. The variant calls corresponded to 270 unique variants of which the majority of 141 (52%) variants were detected by all five institutions, 59 (22%) unique variants were detected by two to four institutions, and 70 (26%) variants by a single institution (Fig. 2). We carried out an in-depth analysis of the discrepant variant calls by separately analyzing the variants that were detected by two to four institutions but missed by the remaining institutions (potentially false-negatives), and the variants detected by only a single institution (potentially false-positives).

Altogether, 59 unique variants were found by some but not all institutions (23 variants were found by four, 21 variants by three, and 15 variants by two institutions), corresponding to 110 missed

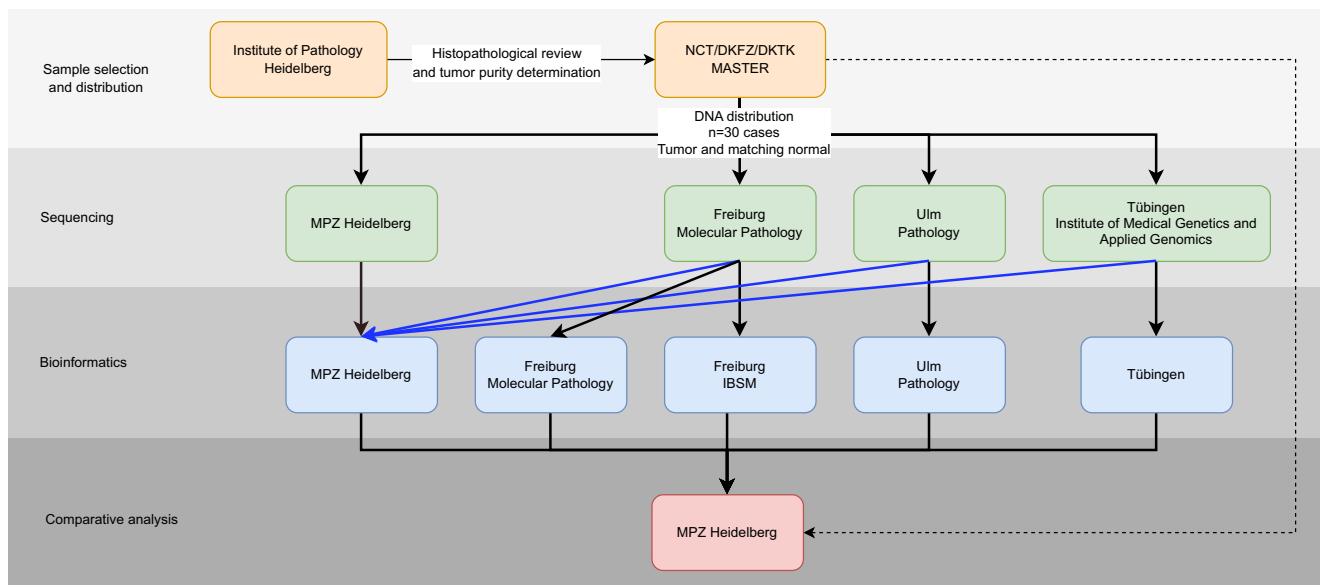


Fig. 1 Overview on the study design. Tumor and normal DNA of 30 patients were distributed by DKFZ and NCT Heidelberg to four participating NGS laboratories. Bioinformatic analysis was performed at five participating departments. Additionally, all sequencing data were collected and analyzed with the same bioinformatic pipeline (blue arrows). Therapeutic relevant results from the DKFZ/NCT/DKTK MASTER program were collected and included in the comparative analysis.

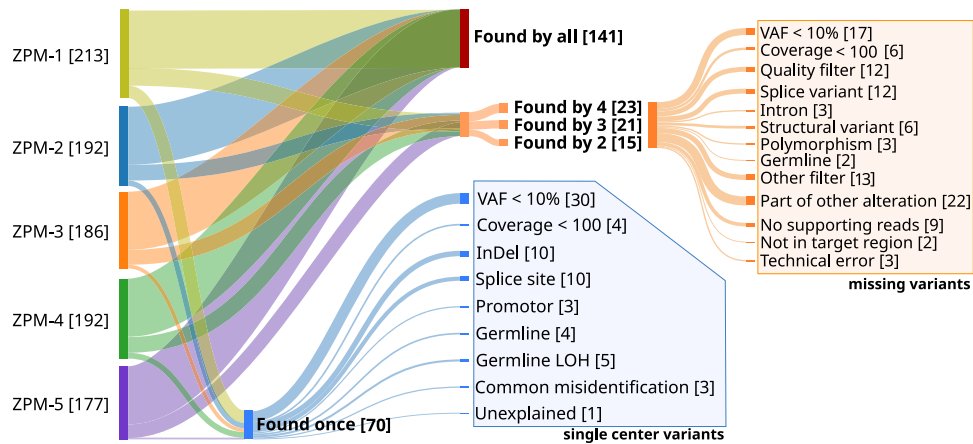


Fig. 2 Inter-institution concordance of the detected somatic variants. Numbers in brackets refer to the numbers of somatic variants. Most variants were detected by all institutions (52%, red bar), while lower percentages of variants were detected by two to four institutions (22%, orange bar), and by only a single institution (26%, blue bar). Potential causes for discordant variant detection were analyzed and the variant sets were split accordingly.

findings of which 23 (21%) had a VAF < 10% or coverage < 100 (Fig. 2, orange subtree). For the remaining variants most misses were due to different filtering procedures including quality, PASS and other filters (37 missed variants, 34%). In addition, 22 (20%) variants were called, but annotated as part of other alterations. Only a small minority of 14 (13%) variant calls were missed due to the lack of supporting reads, technical errors, or not being covered by the target region. Two variant calls (2%) were falsely labeled as germline variant. Six (5%) variants were only identified as structural variants which were not assessed in this study. The remaining 6 (5%) variants were detected, but differently filtered due to intron location or labeling as polymorphism. The percentage of insertions/deletions (indels) missed by at least one of the centers (33%, CI 20–50%) was higher than the percentage of SNV missed by at least one center (20%, CI 15–26%).

Furthermore, 70 variants were called by only a single institution (Fig. 2, blue subtree). Out of these, 30 (43%) variants had VAF < 10% and 4 (6%) had a coverage < 100. Ten (14%) variants were splice site mutations that were detected by other institutions but classified as not affecting alternative splicing. Also, there were 10 (14%) insertions/deletions that were detected by only a single institution. The remaining variants were found in promoter regions (3 variants), were false-positives due to missed germline variants (4 variants), or a special case of loss of heterozygosity (LOH) as a somatic event (5 variants). Three variants were reported in the literature as common misidentified variants due to highly homologous genomic regions⁴³, while only a single reported variant had no supporting reads in the data of the other institutions. The percentage of indels detected by only a single center (24%, CI 12–39%) was similar to the percentage of SNV detected by a single center (29%, CI 23–35%).

Unfortunately, there is no ground truth that can serve as absolute reference for variant calling in clinical tissue samples. To address this limitation, the performance of the participating institutions was evaluated with respect to three different references: (i) a consensus list including all variants detected by at least three institutions (consensus 3x), and (ii) a consensus list including all variants detected by at least two institutions (consensus 2x), and (iii) the TSO500 assay that is used at our institution as validated laboratory developed test (LDT) for clinical mutation testing (Fig. 3a). When using the three-institution consensus as reference, all institutions achieved 100% sensitivity for 9 and at least 85% sensitivity for 16 of the 30 analyzed cases. Across the analyzed 30 cases, the five institutions achieved a positive percentage agreement (PPA) of 91–95% for the three-institution reference, 87–91% for the two-institution reference,

and 90–94% for the TSO500 data reference (Fig. 3b). The corresponding PPV was in the range of 82–95%, 85–98%, and 78–86%, respectively (Fig. 3b).

Therapeutically relevant variants

We carried out an in-depth analysis of variants in druggable genes according to OncoKB, evidence level 1–4⁴⁴. Of 17 variants in druggable genes 16 were detected by all five institutions (Fig. 3c). The *NF-1* variant in sample 26 was reported by three institutions, while two institutions identified it as structural variant. The *ATM* variant in sample 14 was only detected by a single institution. We also compared the detected variants to the therapeutically relevant variants reported in the DKFZ/NCT/DTKK MASTER program (Supplementary Figure 2). The vast majority including 34 of 36 (94%) variants was detected by all institutions and the remaining two variants were found by all but one institution.

Calling of somatic copy number alterations

For each of the 30 tumors, segmentations of the genome in regions of constant copy number (CN) were compared pairwise between institutions. The genomic regions were classified as (i) exact matches, (ii) regions matching after considering genome duplication calls, and (iii) non-matching regions (Fig. 4). Summarizing the results for all 30 tumors, 76% of the genomic regions were matching exactly or after considering genome duplication calls (Fig. 4a, green/purple bars), while copy numbers were non-matching for 24% of the genomic regions (Fig. 4a, red bars). CN differences in divergent intervals were predominantly off by a CN of 1 (56%), followed by CN of 2 (26%), CN of 3 (11%), and higher CN (8%) (Supplementary Figure 3).

To distinguish between wet-lab variation and bioinformatic variability, we reanalyzed the data from all sequencing institutions using the same bioinformatic pipeline. Now 77% of the copy numbers were matching exactly or after considering genome duplications, while the remaining 23% of the copy number were non-matching (Supplementary Fig. 4). Thus, the result did hardly improve when harmonizing the bioinformatics and we can conclude that wet-lab variability was the main contributor to the observed inter-institution CN variability.

Tumor purity (determined by bioinformatics) appeared to be an important confounder for the determination of CNA: Cases with low tumor purity or variable estimates of tumor purity (cases 5, 16, 17, 20) had worse agreements between institutions (Supplementary Fig. 5).

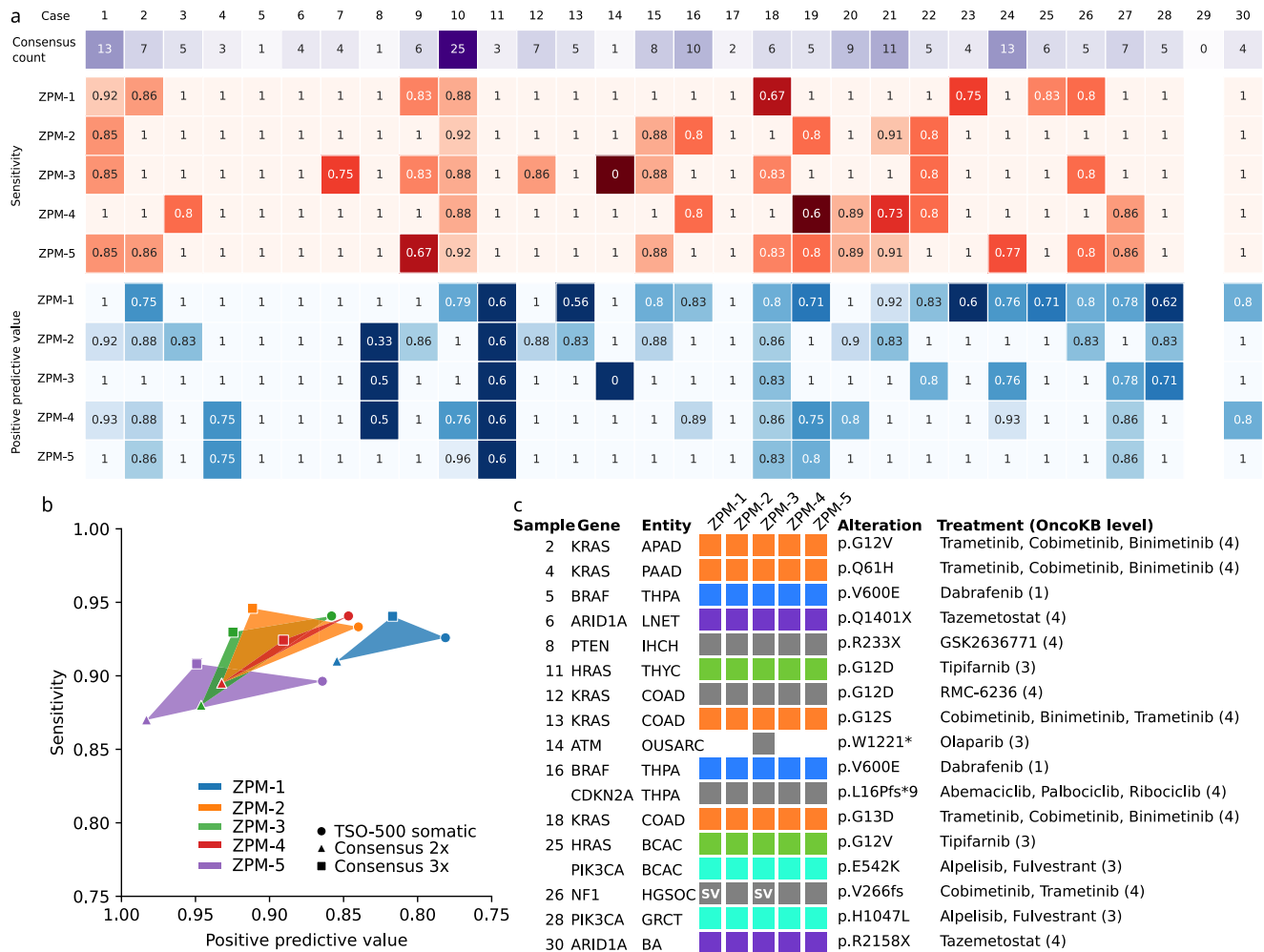


Fig. 3 Comparison of the somatic variants detected by each of the institutions with a consensus list including all variants detected by at least three institutions. **a** Number of variants (3x consensus) for each of the cases as well as sensitivity and positive predictive value for each of the cases and institutions. Empty boxes (case 29) = no variants detected. **b** Sensitivity and positive predictive value for each of five institutions in comparison to three different references: consensus list of variants found in at least two institutions (Consensus 2x, triangle), consensus of variants found in at least three institutions (Consensus 3x, square), and TSO500 (circle). **c** Inter-institution concordance of therapeutic relevant somatic variants (OncoKB levels 1–4) with associated treatments and their OncoKB level. Boxes indicate detected variants and are colored by treatment option, variants marked with “SV” were found as structural variant.

Therapeutic relevant gene amplifications and deletions

We carried out an in-depth analysis of all high-level ($CN \geq 5$ delta from ploidy) amplifications ($n = 22$) and deep biallelic, as well as monoallelic gene deletions ($n = 23$) that were detected by at least two institutions (Fig. 4b). For 16 (73%) of the highly amplified genes, all institutions reported a high-level gene amplification. In five genes only two to four institutions reported a high-level amplification while the others reported an amplification below the high-level threshold. For the remaining single gene, four institutions reported a high-level amplification, while a single institution did not report any amplification. Eleven deletions (48%) were detected by all institutions of which two were deep deletions found by four institutions. Seven deletions were missed by a single institution, while the remaining five deletions were not reported by either two or three institutions. For sample 7, institution ZPM-4 detected deletion of *CDKN2A* and *CDKN2B* with $CN = 1$, while deep deletion of the two genes was detected by all other institutions. This discrepancy persisted when data processing was performed using a uniform bioinformatic pipeline meaning that it was caused by wet-lab variability.

Complex biomarkers

The HRD scores of the 30 tumors correlated strongly between the institutions (Pearson R between 0.79 and 1, Fig. 5). We observed systematic deviations between institutions in the range of -15% to $+9\%$.

Four institutions used the same bioinformatic workflow (Sequenza⁴⁵ and scarHRD⁴⁶) and the same cut-off point of 42 for HRD calling, with one institution using a different segmentation tool (ClinCNV⁴⁷). Different numbers of HRR-deficient tumors were detected: ZPM-4 reported five, ZPM-1 reported eight, ZPM-2 and ZPM-3 each seven, and ZPM-5 ten HRR-deficient cases. For five tumors, all institutions concordantly reported HRD (cases 9, 14, 15, 22, 26). For two additional tumors (cases 21 and 27), four institutions concordantly reported HRD, while the cut-off point was just missed by a single institution (HRDsum=37 and 39, both ZPM-4). In four cases (case 12, 16, 18, and 24) HRD was detected only by a single institution, once by ZPM-1 and three times by ZPM-5.

The TMB scores of the 30 tumors correlated strongly between the participating institutions (Pearson R between 0.97 and 0.99, Fig. 5). In-depth analyses showed a significant systematic deviation of institution ZPM-5 that reported 9–20% lower TMB

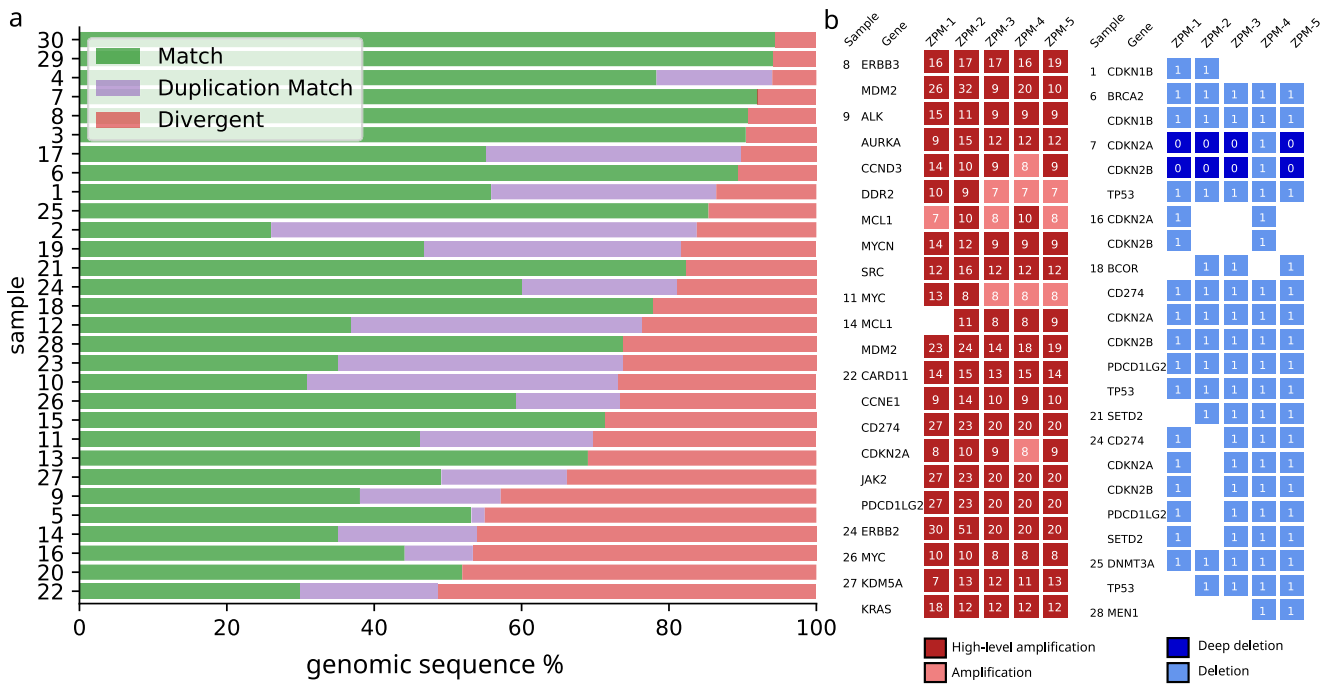


Fig. 4 Inter-institution concordance of CNA calls. a Samples were compared between all pairs of institutions and the genomic sequence was split into segments with concordant CN (green), segments with concordant CN after correction for genome duplication (purple), and discordant CN (red). **b** Inter-institution concordance of amplifications and deletions in the set of oncogenic/likely oncogenic (according to OncoKB) genes. Colored box = CNA detected with number referring to the detected CN with amplifications in red and deletions in blue.

scores compared to ZPM-1, ZPM-2, ZPM-3, and ZPM-4. Two of the tumors (cases 10 and 15) were identified as TMB-high by all institutions. Both were also reported as TMB-high in the DKFZ/NCT/DKTK MASTER program. Case 1 was reported as TMB-high by a single institution (TMB = 10.19) and close to the threshold of 10 mut/MB by the remaining four institutions (TMB between: 9.35–9.95). The remaining 27 tumors were concordantly reported as TMB-low by all institutions.

All tumors in the study cohort were MSS/MSI-L according to the NCT DKFZ/NCT/DKTK MASTER program. Accordingly, none of the institutions reported MSI-H for any of the tumors (Fig. 5). Four institutions used the same bioinformatic tool MSIsensor-pro⁴⁸ resulting in a high concordance between MSI scores (Pearson R between 0.89 and 1). The usage of the tool Mantis⁴⁹ by ZPM-1 together with a different cut-off point of 40% by ZPM-1 resulted in lower correlation of the MSI scores with the other institutions, but concordance concerning MSI status.

Dissecting wet-lab and bioinformatic variability

To distinguish between wet-lab and bioinformatic variability, (i) sequencing data from three wet-labs were reanalyzed using the same bioinformatic pipeline and (ii) the sequencing data generated in Freiburg were analyzed by the bioinformatic pipelines of the Freiburg Institute of Pathology, the Freiburg ISBM, and the Heidelberg Institute of Pathology (Fig. 1).

Results reveal that the median standard deviation (SD) separated for both wet-lab and bioinformatic effects is of relatively low influence for each biomarker (Fig. 6). For HRD, the median SD was 3.6 for the deviations by different wet-lab procedures and 1.7 for the deviations explained by different bioinformatic, both values being small compared to the cut-off point of 42. For TMB, the median SD was 0.34 mut/MB for the wet-lab and 0.18 mut/MB for the bioinformatics evaluation. Although the variability of TMB attributed to the wet-lab procedures was significantly higher than the bioinformatic variability, the former was still very low compared to the cut-off point of 10 mut/MB. For MSI scores, the

median SD was 0.069% for the inter-wet-lab variability and 0.017% for the bioinformatic variability, both values being small compared to the cut-off point of 10%. For tumor purity, estimations also showed a very low median SD of 0.8% for the wet-lab divergence but showed outliers up to 45%, while the deviations in tumor purity estimation explained by bioinformatic variance had a median SD of 0.6% and outliers of up to 30%. For tumor ploidy, the median SD was similar for the variability explained by wet-lab difference (0.13) and for variability from bioinformatic processing (0.18).

For some of the samples, a high variability was observed simultaneously for several complex biomarkers. Cases 16, 28, and 29 showed high variability of HRDsum and simultaneously of the tumors purity estimations. As shown previously⁵⁰, the influence of tumor purity and ploidy estimation on HRDsum scores calls for an accurate histological determination of tumor purity to improve the selection of a correct HRD score. Case 28 showed a high inter-wet-lab variability simultaneously for HRDsum, TMB, and tumor purity.

DISCUSSION

We performed a multicenter comparison of WES analysis of clinical cancer samples covering the entire sample-to-result workflow. This pilot study aimed to assess the level of concordance as well as to identify factors of inter-institutional variability. To this end, we analyzed (i) the inter-center concordance of the results and (ii) the dry lab as well as wet lab factors influencing the results. The study focused on three clinically relevant data layers, somatic variants, CNAs, and complex biomarker (HRD, TMB, MSI), and revealed high concordance between the participating institutions. However, we also identified areas of variability, notably the identification of CNAs. Collectively, these data provide a solid source and basis for standardization and harmonization of WES as a clinical-grade test and contribute to the conceptualization and design of future external quality assessment (EQA) schemes.

The majority of variants was detected by all institutions (52%) and almost all clinically relevant mutations (16 of 17) were

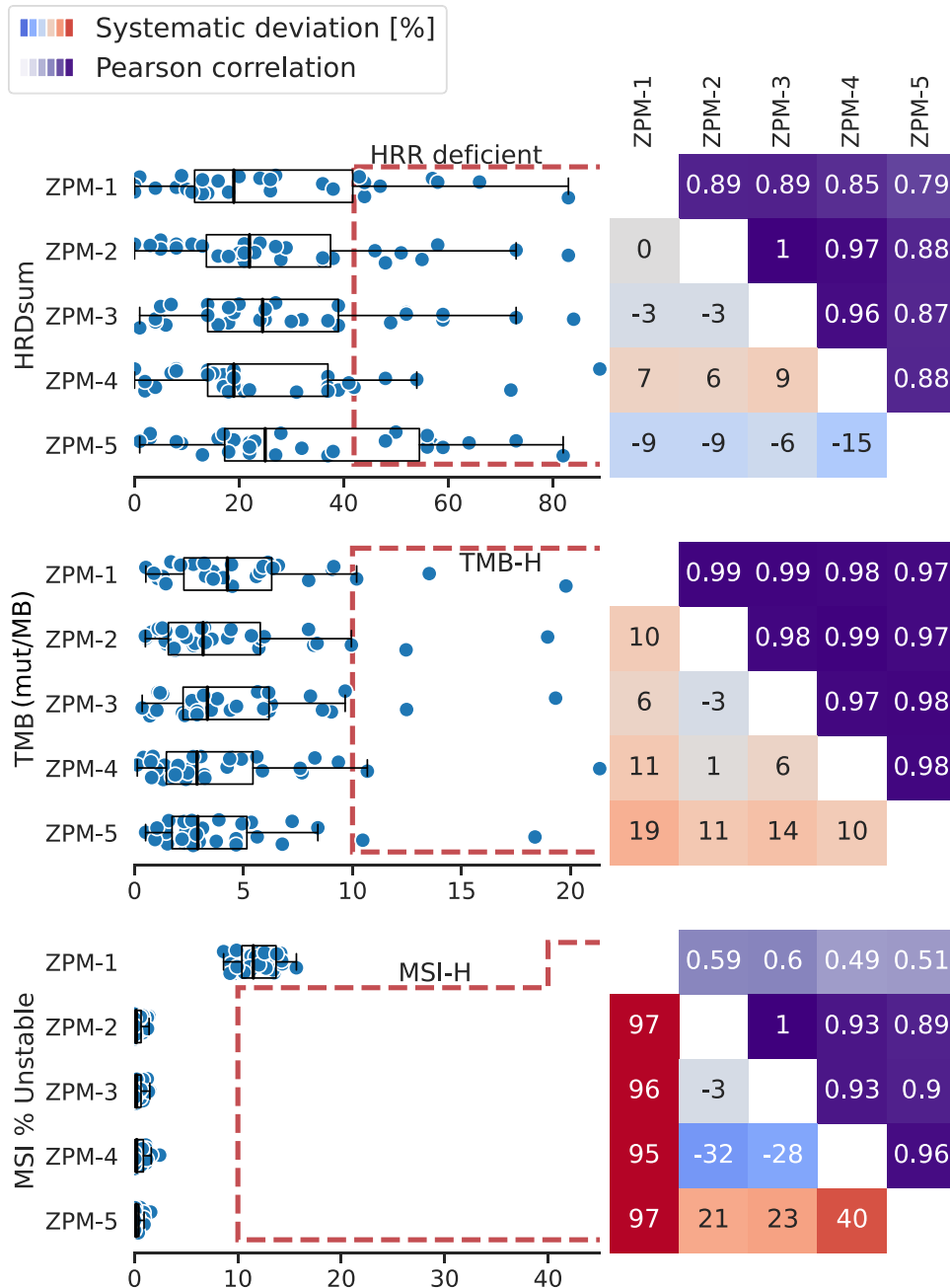


Fig. 5 Inter-institution concordance of HRD, TMB, and MSI scores. Left: Distribution of the complex biomarkers scores together with the institution-specific cut-off points. Boxplots show the median, quartiles and whiskers up to 1.5 times of the box size. Right: Pearson correlation of the scores between pairs of institutions (upper triangle). Systematic deviation (in %) between two institutions (lower triangle). Red = systematic higher scores. Blue = systematic lower scores.

consistently identified by the five participating institutions. In most cases the reasons for discrepancies could be identified, which are predominantly attributable to differences in quality and annotation-based filtering. Only a small fraction (1.5%) of consensus variants were missed due to shortcomings in sequencing or bioinformatic analysis. Our comparison of the WES data to the TSO500 panel derived data also showed that even mutations with lower VAF (< 10%) could be identified in most institutions, yet are the most prominent reason for potential false-positive and false-negative calls. The clinically relevant *ATM* variant found by only one institute also showed low VAF (5.3%) and coverage (78X). In multi-center panel-based somatic variant calling studies detection rates of 80–100% were shown for clinically relevant

alterations^{29–32}. In one of the studies, the detection rate for SNV dropped to 82% when the restriction to clinically relevant variants was lifted³². In contrast to variant caller comparison studies^{26,28} where low concordance was observed, the downstream comparison utilized in our study improved results dramatically, even though a variety of variant callers was utilized (Supplementary Table 2). While detection rates of low frequency somatic variants are presumably biased using WES compared to panels due to reduced coverage, the advantages in routine diagnostics are profound. Alterations found in genes that are currently not clinically relevant can be reanalyzed in the future if required and complex biomarkers can be inferred directly from the same data.

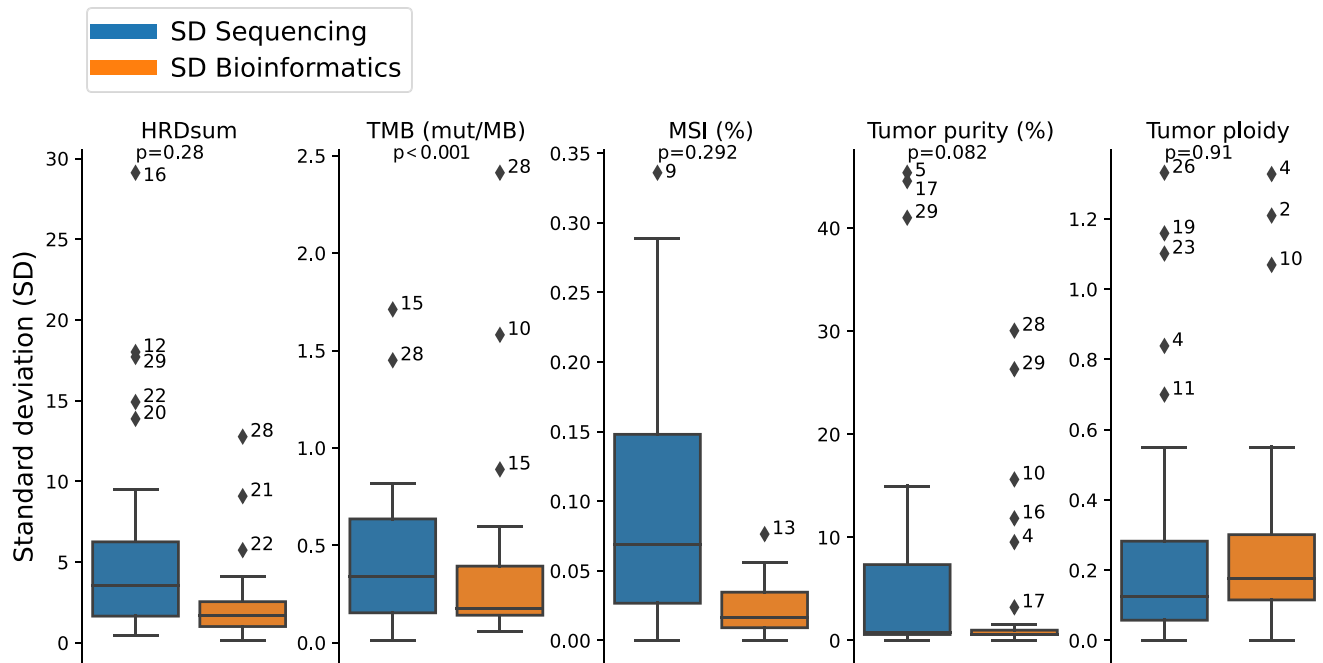


Fig. 6 Comparison of bioinformatic and wet-lab inter-institution variability. To obtain “SD Bioinformatics”, the data of a single sequencing institution were evaluated by three bioinformatic institutions. To obtain “SD Sequencing” the data of three sequencing institutions were evaluated by a single bioinformatic institution. Outliers are annotated with the sample number. Boxplots show the median, quartiles and whiskers up to 1.5 times of the box size.

Analysis stratified by variant type across the five participating centers showed that variants were missed more often for indels (33%) than for SNV (20%), while false positive variant detections were similar for both indels and SNV. These results are in line with earlier studies showing lower agreement of different variant callers for indels compared to SNV²⁴. Thus, attention should be paid to implementation of a sensitive indel calling in bioinformatic pipelines to avoid missing of clinically relevant alterations such as for example activating indels in the exon 19 of the EGFR gene or truncating indels in tumor suppressor genes.

Compared to point mutations a larger variability was observed for CNA calling. High-level gene amplifications were found concordantly for most high-level amplifications calculated as delta from the ploidy. This raises the question as to which approach should be utilized. Gene deletions were shown to be disparate for several cases as a result of the previously described wet-lab variability. There is a lack of multi-center studies on CNA calling variability, but some reports indicating that the use of different bioinformatic tools for CNA detection is associated with significant variability^{34,35}. However, we observed a high concordance of oncogenic/likely oncogenic amplifications and deletions between the utilized tools Sequenza (ZPM-2 to ZPM-5) and ClinCNV (ZPM-1) (Fig. 4b).

For the complex biomarkers HRD, TMB, and MSI, strong correlations of scores and high concordance of the diagnostic classification were observed even though bioinformatic pipelines were not harmonized prior to the pilot study. In contrast to previous studies performed using gene-panels, a higher concordance was observed for TMB^{39,51,52}, which is attributable to the broader genomic footprint of WES. The high concordance of WES TMB values was also reported previously with correlations between 0.85 and 0.99 depending on tumor entity²⁰. The same bioinformatic tools were used by all institutions for the determination of the HRD score. This leaves the observed variability attributed to wet-lab variance and configuration of bioinformatic tools. No multi-center study comparing HRD calls is currently known to the authors thus precluding consideration of

the deviations in this context. Separation of bioinformatic and wet-lab variability allowed for an in-depth examination of inter-center variability leading to several lessons learned which will be discussed below.

Since the study focused on the entire sample-to-result workflow (result quality) including the laboratory analysis and the bioinformatic pipeline, we did not standardize wet-lab methods prior to the study (such as pre-analytics, library preparation, and sequencing). The study also did not specify the use of certain alignment or variant calling tools. This approach was taken intentionally to compare the performance of different NGS labs in the determination of clinically relevant somatic variants and genomic biomarkers in a real-world setting but is also a limitation of the study. A systematic comparison of different wet-lab methods, different variant callers or other bioinformatic tools was not intended and not feasible based on acquired data.

The study lacked a ground truth for the somatic variants, the CNA and complex biomarkers that could have been used as reference for comparison with the results of the participating institutions. Absence of a ground truth is unavoidable in comprehensive analysis of clinical tissue samples where a large number of genes are interrogated, but this is addressed in the current study by using a consensus reference and by comparison with deeper coverage panel sequencing data. Other studies have evaluated the performance of WES and WGS in samples with available ground truth but had to use artificial models such as cell lines^{26,33} or sequencing data with artificially added mutations to achieve this²⁵, which does not reflect the clinical diagnostic setting. Taken together, the multitude of genetic events analyzed in WES coupled with the absence of a simply determinable ground truth when analyzing real-world clinical samples leads to a highly complex exercise compared to single-gene or few-gene tests which needs to be addressed conceptually by new kinds of proficiency tests.

The current study was performed using DNA extracted from fresh-frozen samples. Similar studies are warranted to evaluate the inter-center variability when analyzing formalin-fixed, paraffin-

embedded (FFPE) samples, as low-confidence calls are more difficult to identify correctly in FFPE-material⁵³. A further limiting factor for the generalization of the results was the location of all participating centers in Germany.

Based on our findings the following points should be considered for the implementation of WES as a clinical test in routine diagnostics: (i) Current clinically relevant alterations are identifiable with high precision and sensitivity despite non-harmonized wet lab and dry lab procedures. (ii) Detected somatic SNVs and indels show fairly high concordance with divergence mainly observed for low VAF variants in low-coverage regions, which were the main source of decreased sensitivity and PPV. Moreover, differences in the assessed region, annotation of variants in splice sites and promoter, and other quality or annotation-based filter criteria led to a small number of differentially reported variants, indicating the need for further harmonization of bioinformatics pipelines. (iii) Complex biomarker results show high concordance but are sensitive to issues with purity and ploidy estimation. (iv) High-level gene amplifications were reliably identified but other CNA (low level amplifications and deletions) were more difficult to detect primarily due to wet-lab variance and this needs to be reviewed accordingly.

In summary, calling of somatic variants was highly concordant with a PPA of 91–95% and a PPV of 82–95% compared against the three-institution consensus and full agreement for 16 of 17 druggable targets. Complex biomarkers correlated strongly between institutions (HRD: 0.79–1.00, TMB: 0.97–0.99) and all institutions agreed upon microsatellite stability status. Our data argue for the development of stringent standards defining a clinical-grade WES test and emphasize the need for harmonization of both wet lab and dry lab settings between institutions to ensure robust and comparable diagnostic results. The design of the current study and the strategies developed for the data evaluation will serve as a basis for a WES pilot involving 20 centers within the German Network for Personalized Medicine (DNPM⁵⁴) and informs the development of future EQA schemes (e.g. in GenomDE⁵⁵), including ring trials, for clinical-grade WES tests.

METHODS

Case selection and distribution

Paired tumor and normal DNA samples from 30 patients were subjected to WES at four NGS laboratories and analyzed at five bioinformatic departments of the ZPM (Fig. 1). The 30 cases were selected from patients enrolled in the German DKFZ/NCT/DKTK MASTER program to represent a wide variety of molecular alterations. DKFZ/NCT/DKTK MASTER (NCT05852522) is a multi-center registry trial for prospective, biology-driven stratification of younger adults with advanced-stage cancer across all entities and patients with rare tumors. MASTER patients consented to banking of tumor and control tissue, molecular profiling of both samples, and clinical data collection (S-206/2011, Ethics Committee of the Medical Faculty of Heidelberg University). Tumor and germline DNA were shipped to each of the participating centers based on specific material transfer agreements. The cohort included a variety of rare cancer types including carcinomas, sarcomas, and other tumor entities, with a histopathologically determined tumor purity between 46% and 90% (Supplementary Table 1).

Sequencing and data analysis

WES was performed independently by each laboratory using the established protocols with different sequencers, library kits, and sequencing chemistry. ZPM-1 and ZPM-5 generated sequencing libraries for tumor and normal samples using 100 ng of DNA and Twist Human Comprehensive Exome 2.0 (with additional custom regions), which were sequenced on Illumina NovaSeq6000 or NextSeq550Dx (see Supplementary Table 3). ZPM-4 also used the

Twist Human Comprehensive Exome 2.0 sequenced on Illumina NovaSeq6000 with a varying amount of DNA for tumor and normal samples (Supplementary Table 3). ZPM-3 used the Agilent SureSelect XT Human All Exon V8 and sequenced on Illumina NextSeq 550 (Supplementary Table 3). Bioinformatic processing was performed independently in each institution using its custom pipeline (Supplementary Table 2). Quality control metrics were calculated using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and qualimap bamqc⁵⁶ with each center using their target region as reference.

Furthermore, all tumor samples were re-analyzed using the Illumina TruSight Oncology 500 (TSO500)-panel. DNA integrity was assessed using the Genomic DNA ScreenTape Analysis on a 4200 TapeStation System (both Agilent, Santa Clara, California). To fragment DNA to a mean fragment size of around 200 bp, 80 ng DNA of each sample was sheared for 50 to 78 seconds using a focused ultrasonicator ME220 (Covaris, Woburn, Massachusetts). The library preparation for the capture-based TSO500 panel was performed according to the manufacturers' manual with a final amplification of 15 cycles for the final libraries. Quality control and quantification was conducted using the KAPA SYBR Library Quantification Kit on a StepOnePlus quantitative PCR system (both Thermo Fisher Scientific). Up to eight libraries were sequenced simultaneously on a NextSeq 500 (Illumina) using high-output cartridge and v2 chemistry resulting in ~100 M read pairs (2 × 150 bp) per sample. All assays were performed according to the manufacturers' protocols. DNA sequencing data was analyzed by TruSight Oncology 500 Local App (Illumina, pipeline version 2.2).

Data collection and reference gene list

The following data were collected from the participating institutions in a predefined format: (1) variant calls, (2) copy number variation segments including allele-specific information, and (3) TMB, HRD, and MSI scores. The concordance of variant calls was evaluated in a list of 494 genes that included the American College of Medical Genetics (ACMG) list of cancer-related genes⁵⁷ and genes utilized in the NCT MASTER program (Supplementary Table 3). Therapeutically relevant results for targetable genetic alterations reported in the NCT MASTER program were also acquired for each case.

Somatic variants were classified as therapeutically relevant if an annotation in level 1–4 of OncoKB of the corresponding entity or higher-level entity based on OncoTree⁵⁸ was found.

Oncogenic deletions, oncogenic and likely oncogenic amplifications were downloaded from OncoKB⁴⁴, and intersected with the segmented CNAs using BedTools⁵⁹. Reported deep deletions and high-level amplifications (CN ≥ 5) were filtered based on the OncoKB list.

Variant calls from the TSO500 panel were filtered by the intersection of genes present in the gene list and panel ($n = 274$). As the TSO500 panel was used for sequencing of tumor DNA, germline variants were removed by subtracting the WES germline calls from all institutions combined.

Bioinformatic versus sequencing variance

To distinguish between wet-lab and bioinformatic variability, raw sequencing data (FASTQ files) from all sequencing centers were re-analyzed using the same bioinformatic pipeline (Fig. 1, blue arrows). Conversely, raw sequencing data from the Freiburg sequencing center were analyzed using three different bioinformatic pipelines. As a result, wet-lab variability and bioinformatic variability could be evaluated separately. For the complex biomarkers, values were normalized by fitting a linear model with the intercept fixed at zero against the Heidelberg pipeline prior to calculation of the standard deviations. Statistical difference was

assessed using the paired Wilcoxon test. P -values < 0.05 were considered significant.

Inter-institution comparison of somatic variant, copy number, and complex biomarkers

Somatic variants were compared based on their chromosomal position, and alteration at the DNA level. Percentages of SNV and of indels were reported with 95% confidence intervals and calculated using the Clopper-Pearson method.

Copy numbers (CN) were collected as genomic intervals with position and respective allele-specific copy numbers. Regions without reported CN change were set to a CN of two. The proportions of the genome having CN = 0, 1, 2, 3, 4, ... and the predominant ploidy were extracted. CN for each base of the genome were compared between each pair of two institutions. CN agreement was assessed as follows: (1) exact match, (2) discrepancy explained by the difference of the predominant ploidies (e.g., one institution reported a genome duplication, while the other institution did not), and (3) unexplained discrepancy.

HRD, TMB, and MSI scores were compared pairwise between institutions by correlations analysis (Pearson correlation) and by fitting of a linear model with the intercept fixed at zero. Statistical significance of difference was assessed with a one-sided Wilcoxon test and $p < 0.05$. All institutions used a cut-off point of 42 for HRD and a cut-off point of 10 mutations per megabase (MB) for TMB. For MSI, four institutions used a cut-off point of 10%, while a single institution used a cut-off point of 40% due to the application of a different bioinformatic approach for MSI detection (Supplementary Table 2). For sample 14 institution ZPM-1 was unable to determine the tumor purity and therefore correlations were calculated using 29 samples.

Software for statistical analysis and graphics generation

Statistical analysis and figure creation were performed using Python with `scipy`⁶⁰, `pybedtools`⁶¹, and `numpy`⁶². Figures were created using `Matplotlib`⁶³, `seaborn`⁶⁴, `SankeyMatic` (<https://sankeymatic.com/>), and `Pandas`⁶⁵.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The sequencing data generated in the study are available from the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega/datasets>) under the Accession Number EGAD00001011087.

CODE AVAILABILITY

The code used to compare the results of the five participating institutions is available upon request to the corresponding authors.

Received: 3 April 2023; Accepted: 26 September 2023;

Published online: 20 October 2023

REFERENCES

- Mateo, J. et al. Delivering precision oncology to patients with cancer. *Nat. Med.* **28**, 658–665 (2022).
- Tsimberidou, A. M., Fountzilias, E., Nikanjam, M. & Kurzrock, R. Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer Treat. Rev.* **86**, 102019 (2020).
- Koboldt, D. C. Best practices for variant calling in clinical sequencing. *Genome Med.* **12**, 91 (2020).

- Mosele, F. et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. *Ann. Oncol.* **31**, 1491–1505 (2020).
- Jennings, L. J. et al. Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagn.* **19**, 341–365 (2017).
- Bewicke-Copley, F., Kumar, E. A., Palladino, G., Korfi, K. & Wang, J. Applications and analysis of targeted genomic sequencing in cancer studies. *Comput. Struct. Biotechnol. J.* **17**, 1348 (2019).
- Robinson, D. R. et al. Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).
- Priestley, P. et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).
- Horak, P. et al. Comprehensive genomic and transcriptomic analysis for guiding therapeutic decisions in patients with rare cancers. *Cancer Discov.* **11**, 2780–2795 (2021).
- Schipper, L. J. et al. Clinical impact of prospective whole genome sequencing in sarcoma patients. *Cancers* **14**, 436 (2022).
- Schipper, L. J. et al. Complete genomic characterization in patients with cancer of unknown primary origin in routine diagnostics. *ESMO Open* **7**, 100611 (2022).
- Samsom, K. G. et al. Feasibility of whole-genome sequencing-based tumor diagnostics in routine pathology practice. *J. Pathol.* **258**, 179–188 (2022).
- Hoes, L. R. et al. Patients with Rare Cancers in the Drug Rediscovery Protocol (DRUP) Benefit from Genomics-Guided Treatment. *Clin. Cancer Res.* **28**, 1402–1411 (2022).
- Van Allen, E. M. et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688 (2014).
- Niguidula, N. et al. Clinical whole-exome sequencing results impact medical management. *Mol. Genet. Genom. Med.* **6**, 1068–1078 (2018).
- Dotolo, S. et al. Bioinformatics: From NGS Data to Biological Complexity in Variant Detection and Oncological Clinical Practice. *Biomedicines* **10**, 2074 (2022).
- Chan, T. A. et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol. J. Eur. Soc. Med. Oncol.* **30**, 44–56 (2019).
- Rempel, E. et al. Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (HRD). *npj Precis. Oncol.* **6**, 1–13 (2022).
- Yu, F., Makrigiorgos, A., Leong, K. W. & Makrigiorgos, G. M. Sensitive detection of microsatellite instability in tissues and liquid biopsies: Recent developments and updates. *Comput. Struct. Biotechnol. J.* **19**, 4931–4940 (2021).
- Ramaraio-Milne, P. et al. Comparison of actionable events detected in cancer genomes by whole-genome sequencing, in silico whole-exome and mutation panels. *ESMO Open* **7**, 100540 (2022).
- Auzanneau, C. et al. Feasibility of high-throughput sequencing in clinical routine cancer care: lessons from the cancer pilot project of the France Genomic Medicine 2025 plan. *ESMO Open* **5**, e000744 (2020).
- Massard, C. et al. High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: results of the MOSCATO 01 trial. *Cancer Discov.* **7**, 586–595 (2017).
- Cai, L., Yuan, W., Zhang, Z., He, L. & Chou, K.-C. In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Sci. Rep.* **6**, 36540 (2016).
- Kroigård, A. B., Thomassen, M., Lænholm, A.-V., Kruse, T. A. & Larsen, M. J. Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. *PLOS ONE* **11**, 1–15 (2016).
- Ewing, A. D. et al. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat. Methods* **12**, 623–630 (2015).
- Zhao, Y. et al. Whole genome and exome sequencing reference datasets from a multi-center and cross-platform benchmark study. *Sci. Data* **8**, 296 (2021).
- Xiao, W. et al. Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing. *Nat. Biotechnol.* **39**, 1141–1150 (2021).
- Barbitoff, Y. A., Abasov, R., Tvorogova, V. E., Glotov, A. S. & Predeus, A. V. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. *BMC Genomics* **23**, 155 (2022).
- Spence, T. et al. Inter-laboratory proficiency testing scheme for tumour next-generation sequencing in Ontario: a pilot study. *Curr. Oncol. Tor. Ont.* **26**, e717–e732 (2019).
- Quy, P. N. et al. Inter-assay variability of next-generation sequencing-based gene panels. *BMC Med. Genomics* **15**, 86 (2022).
- Maekawa, M. et al. Precision cancer genome testing needs proficiency testing involving all stakeholders. *Sci. Rep.* **12**, 1494 (2022).

32. Davies, K. D. et al. Multi-Institutional FASTQ File Exchange as a Means of Proficiency Testing for Next-Generation Sequencing Bioinformatics and Variant Interpretation. *J. Mol. Diagn.* **18**, 572–579 (2016).
33. Karimnezhad, A. et al. Accuracy and reproducibility of somatic point mutation calling in clinical-type targeted sequencing data. *BMC Med. Genomics* **13**, 156 (2020).
34. Gabriellaite, M. et al. A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. *Cancers* **13**, 6283 (2021).
35. Zare, F., Dow, M., Monteleone, N., Hosny, A. & Nabavi, S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinforma.* **18**, 286 (2017).
36. Nam, J.-Y. et al. Evaluation of somatic copy number estimation tools for whole-exome sequencing data. *Brief. Bioinform.* **17**, 185–192 (2016).
37. Kadalayil, L. et al. Exome sequence read depth methods for identifying copy number changes. *Brief. Bioinform.* **16**, 380–392 (2015).
38. Merino, D. M. et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J. Immunother. Cancer* **8**, e000147 (2020).
39. Vega, D. M. et al. Aligning tumor mutational burden (TMB) quantification across diagnostic platforms: phase II of the Friends of Cancer Research TMB Harmonization Project. *Ann. Oncol. J. Eur. Soc. Med. Oncol.* **32**, 1626–1636 (2021).
40. Lambin, S. et al. 33P - Tumour mutational burden ring trial: Evaluation of targeted next-generation sequencing platforms for implementation in clinical practice. *Abstr. Book ESMO Immuno-Oncol. Congr. 2019 11–14 Dec. 2019 Geneva Switz.* **30**, xi10 (2019).
41. Velasco, A. et al. Multi-center real-world comparison of the fully automated Idylla™ microsatellite instability assay with routine molecular methods and immunohistochemistry on formalin-fixed paraffin-embedded tissue of colorectal cancer. *Virchows Arch. Int. J. Pathol.* **478**, 851–863 (2021).
42. Stenzinger, A. et al. Trailblazing precision medicine in Europe: A joint view by Genomic Medicine Sweden and the Centers for Personalized Medicine, ZPM, in Germany. *Precis. Med. Cancer* **84**, 242–254 (2022).
43. Bowler, T. G. et al. Misidentification of MLL3 and other mutations in cancer due to highly homologous genomic regions. *Leuk. Lymphoma* **60**, 3132–3137 (2019).
44. Chakravarty, D. et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* 1–16. <https://doi.org/10.1200/PO.17.00011> (2017).
45. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol. J. Eur. Soc. Med. Oncol.* **26**, 64–70 (2015).
46. Sztupinszki, Z. et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *Npj Breast Cancer* **4**, 16 (2018).
47. Demidov, G., Sturm, M. & Ossowski, S. ClinCNV: multi-sample germline CNV detection in NGS data. *bioRxiv* 2022.06.10.495642. <https://doi.org/10.1101/2022.06.10.495642> (2022).
48. Jia, P. et al. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. *Genomics Proteom. Bioinforma.* **18**, 65–71 (2020).
49. Kautto, E. A. et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget* **8**, 7452–7463 (2017).
50. Menzel, M. et al. Accurate tumor purity determination is critical for the analysis of homologous recombination deficiency (HRD). *Transl. Oncol.* **35**, 101706 (2023).
51. Budczies, J. et al. Optimizing panel-based tumor mutational burden (TMB) measurement. *Ann. Oncol. J. Eur. Soc. Med. Oncol.* **30**, 1496–1506 (2019).
52. Stenzinger, A. et al. Harmonization and Standardization of Panel-Based Tumor Mutational Burden Measurement: Real-World Results and Recommendations of the Quality in Pathology Study. *J. Thorac. Oncol.* **15**, 1177–1189 (2020).
53. Oh, E. et al. Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples. *PLoS One* **10**, e0144162 (2015).
54. Illert, A. L. et al. The German Network for Personalized Medicine to enhance patient care and translational research. *Nat. Med.* **29**, 1298–1301 (2023).
55. GenomDE <https://genom.de/> (2023).
56. Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **32**, 292–294 (2016).
57. Miller, D. T. et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **24**, 1407–1414 (2022).
58. Kundra, R. et al. OncoTree: A Cancer Classification System for Precision Oncology. *JCO Clin. Cancer Inf.* **5**, 221–230 (2021).
59. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
60. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
61. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
62. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
63. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
64. Waskom, M. L. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
65. team, T. pandas development. *pandas-dev/pandas: Pandas*. (Zenodo, 2020). <https://doi.org/10.5281/zenodo.3509134>.

ACKNOWLEDGEMENTS

This work was supported by funds from the state of Baden-Wuerttemberg within the Centers for Personalized Medicine Baden-Wuerttemberg (ZPM). We thank Twist Bioscience for the technical support in the establishing of the Whole Exome Sequencing workflow in the ZPM Pilot. We thank Agilent and Twist for partial provision of sequencing reagents.

AUTHOR CONTRIBUTIONS

M.M.: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing-original draft, and Writing - review & editing. S.O.: Funding acquisition, Supervision, and Writing - review & editing. S.K.: Data curation, Investigation, and Writing-review & editing. P.M.: Data curation, Investigation, and Writing-review & editing. P.H.: Funding acquisition, Resources, Supervision, and Writing-review & editing. R.M.: Funding acquisition, Supervision, and Writing - review & editing. M.B.: Funding acquisition, Supervision, and Writing - review & editing. S.W.: Data curation, Investigation, and Writing-review & editing. M.B.: Data curation, Investigation, and Writing-review & editing. O.N.: Data curation, Investigation, and Writing-review & editing. S.A.-E.: Data curation, Investigation, and Writing - review & editing. C.S.: Data curation, Investigation, and Writing - review & editing. U.M.: Data curation, Investigation, and Writing-review & editing. H.G.: Data curation, Investigation, and Writing - review & editing. V.S.: Data curation, Investigation, and Writing - review & editing. A.F.: Data curation, Investigation, and Writing-review & editing. M.A.: Data curation, Investigation, and Writing - review & editing. T.E.: Data curation, Investigation, and Writing-review & editing. J.N.: Data curation, Investigation, and Writing-review & editing. A.B.: Data curation, Investigation, and Writing-review & editing. C.P.: Project administration and Writing - review & editing. T.H.: Data curation, Investigation, and Writing-review & editing. K.T.: Data curation and Writing - review & editing. O.K.: Data curation, Investigation, and Writing - review & editing. T.P.: Data curation, Investigation, and Writing-review & editing. M.K.: Data curation, Investigation, and Writing - review & editing. K.K.: Data curation, Investigation, and Writing - review & editing. A.O.: Data curation, Investigation, and Writing-review & editing. J.A.: Data curation, Investigation, and Writing - review & editing. A.G.: Data curation, Investigation, and Writing-review & editing. S.L.: Funding acquisition, Supervision, and Writing - review & editing. H.K.: Supervision and Writing - review & editing. F.F.: Supervision and Writing - review & editing. A.I.: Supervision and Writing - review & editing. M.W.: Supervision and Writing - review & editing. P.M.: Supervision and Writing - review & editing. T.S.: Supervision and Writing - review & editing. N.M.: Supervision and Writing - review & editing. P.S.: Supervision and Writing - review & editing. S.F.: Conceptualization, Funding acquisition, Resources, and Writing - review & editing. D.K.: Data curation, Investigation, Methodology, Supervision, and Writing - review & editing. J.B.: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing-original draft, and Writing-review & editing. A.S.: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing-original draft, and Writing - review & editing.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

P.H. reports Consulting or Advisory Role: Platomics GmbH, Honoraria: Roche Pharma AG, Trillium GmbH. R.M. reports participation in Advisory Boards of Roche, AstraZenca; research grants from Bristol-Myers-Squibb. C.S. reports an institutional grant from Illumina and research grants from BMS Stiftung Immunonkologie outside the submitted work. M.K.: speaker's honoraria and travel grants from Veracyte Inc. P.S. reports personal fees for speaker honoraria from Incyte, Roche, Janssen, Novartis, AstraZeneca, Eisai, Leica, grants from Novartis, BMS, AstraZeneca, Illumina, Incyte, and

boards from Incyte, Roche, AstraZeneca, BMS, MSD, Amgen, Janssen, Novartis, Bayer, Eisai, outside the submitted work. S.F. reports consulting or advisory board membership: Bayer, Illumina, Roche; honoraria: Amgen, Eli Lilly, PharmaMar, Roche; research funding: AstraZeneca, Pfizer, PharmaMar, Roche; travel or accommodation expenses: Amgen, Eli Lilly, Illumina, PharmaMar, Roche. D.K. reports personal fees for speaker honoraria from AstraZeneca, and Pfizer, personal fees for Advisory Board from Bristol-Myers Squibb, outside the submitted work. S.L. reports research grant from BMS, advisory board/speaker invitation from AstraZeneca, Eli Lilly, Roche and Takeda outside of this work. A.I.: speaker's honoraria/ advisory boards Takeda, Roche, Amgen, Janssen, Abbvie, Bayer, Incyte, FiDO outside the submitted work. J.B. reports grants from German Cancer Aid and consulting from MSD, outside the submitted work. A.S. reports grants and personal fees from Bayer, BMS, grants from Chugai and personal fees from Astra Zeneca, MSD, Takeda, Seattle Genetics, Novartis, Illumina, Thermo Fisher, Eli Lilly, Takeda, outside the submitted work. A.S. reports participation in Advisory Board/Speaker's Bureau for Astra Zeneca, AGCT, Bayer, Bristol-Myers Squibb, Eli Lilly, Illumina, Janssen, MSD, Novartis, Pfizer, Roche, Seattle Genetics, Takeda, and Thermo Fisher, grants from Bayer, Bristol-Myers Squibb, and Chugai, outside the submitted work. M.M., S.O., S.L., P.M., P.H., M.B., S.W., M.B., O.N., S.A., U.M., H.G., V.S., A.F., M.A., T.E., J.N., A.B., C.P., T.B.H., T.K., O.K., T.P., K.K., A.O., J.A., A.G., S.K., H.K., F.F., A.L., M.W., P.M., T.S., N.M., S.F. report no conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41698-023-00457-x>.

Correspondence and requests for materials should be addressed to Jan Budczies or Albrecht Stenzinger.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023