Journal Pre-proof

Benchmarking Whole Exome Sequencing in the German Network for Personalized Medicine

Michael Menzel, Mihaela Martis-Thiele, Hannah Goldschmid, Alexander Ott, Eva Romanovsky, Janna Siemanowski-Hrach, Lancelot Seillier, Nadina Ortiz Brüchle, Angela Maurer, Kjong-Van Lehmann, Matthias Begemann, Miriam Elbracht, Robert Meyer, Sebastian Dintner, Rainer Claus, Jan Meier-Kolthoff, Eric Blanc, Markus Möbs, Maria Joosten, Manuela Benary, Patrick Basitta, Florian Hölscher, Verena Tischler, Thomas Groß, Oliver Kutz, Rebecca Prause, Doreen William, Kai Horny, Wolfgang Goering, Sugirthan Sivalingam, Arndt Borkhardt, Cornelia Blank, Stefanie Junk, Layal Yasin, Evgeny A. Moskalev, Maria Giulia Carta, Fulvia Ferrazzi, Lars Tögel, Steffen Wolter, Eugen Adam, Uta Matysiak, Tessa Rosenthal, Jürgen Dönitz, Ulrich Lehmann, Gunnar Schmidt, Stephan Bartels, Winfried Hofmann, Steffen Hirsch, Nicola Dikow, Kirsten Göbel, Rouzbeh Banan, Stefan Hamelmann, Annette Fink, Markus Ball, Olaf Neumann, Jan Rehker, Michael Kloth, Justin Murtagh, Nils Hartmann, Phillip Jurmeister, Andreas Mock, Jörg Kumbrink, Andreas Jung, Eva-Maria Mayr, Anne Jacob, Marcel Trautmann, Santina Kirmse, Kim Falkenberg, Christian Ruckert, Daniela Hirsch, Alexander Immel, Wolfgang Dietmaier, Tobias Haack, Ralf Marienfeld, Axel Fürstberger, Jakob Niewöhner, Uwe Gerstenmaier, Timo Eberhardt, Phillip Greif, Silke Appenzeller, Katja Maurus, Julia Doll, Yvonne Jelting, Danny Jonigk, Bruno Märkl, Dieter Beule, David Horst, Anna-Lena Wulf, Daniela Aust, Martin Werner, Kirsten Reuter-Jessen, Philipp Ströbel, Bernd Auber, Felix Sahm, Sabine Merkelbach-Bruse, Udo Siebolts, Wilfried Roth, Silke Lassmann, Frederick Klauschen, Nadine T. Gaisa, Wilko Weichert, Matthias Evert, Sorin Armeanu-Ebinger, Stephan Ossowski, Christopher Schroeder, Christian P. Schaaf, Nisar Malek, Peter Schirmacher, Daniel Kazdal, Nicole Pfarr, Jan Budczies, Albrecht Stenzinger

PII: S0959-8049(24)00962-6

DOI: https://doi.org/10.1016/j.ejca.2024.114306

Reference: EJC114306

To appear in: *European Journal of Cancer*

Received date: 20 June 2024
Revised date: 23 August 2024
Accepted date: 23 August 2024

Please cite this article as: Michael Menzel, Mihaela Martis-Thiele, Hannah Goldschmid, Alexander Ott, Eva Romanovsky, Janna Siemanowski-Hrach, Lancelot Seillier, Nadina Ortiz Brüchle, Angela Maurer, Kjong-Van Lehmann, Matthias Begemann, Miriam Elbracht, Robert Meyer, Sebastian Dintner, Rainer Claus, Jan Meier-Kolthoff, Eric Blanc, Markus Möbs, Maria Joosten, Manuela Benary, Patrick Basitta, Florian Hölscher, Verena Tischler, Thomas Groß, Oliver Kutz, Rebecca Prause, Doreen William, Kai Horny, Wolfgang Goering, Sugirthan Sivalingam, Arndt Borkhardt, Cornelia Blank, Stefanie Junk, Layal Yasin, Evgeny A. Moskalev, Maria Giulia Carta, Fulvia Ferrazzi, Lars Tögel, Steffen Wolter, Eugen Adam, Uta Matysiak, Tessa Rosenthal, Jürgen Dönitz, Ulrich Lehmann, Gunnar Schmidt, Stephan Bartels, Winfried Hofmann, Steffen Hirsch, Nicola Dikow, Kirsten Göbel, Rouzbeh Banan, Stefan Hamelmann, Annette Fink, Markus Ball, Olaf Neumann, Jan Rehker, Michael Kloth, Justin Murtagh, Nils Hartmann, Phillip Jurmeister, Andreas Mock, Jörg Kumbrink, Andreas Jung, Eva-Maria Mayr, Anne Jacob, Marcel Trautmann, Santina Kirmse, Kim Falkenberg, Christian Ruckert, Daniela Hirsch, Alexander Immel, Wolfgang Dietmaier, Tobias Haack, Ralf Marienfeld, Axel Fürstberger, Jakob Niewöhner, Uwe Gerstenmaier, Timo Eberhardt, Phillip Greif, Silke Appenzeller, Katja Maurus, Julia Doll, Yvonne Jelting, Danny Jonigk, Bruno Märkl, Dieter Beule, David Horst, Anna-Lena Wulf, Daniela Aust, Martin Werner, Kirsten Reuter-Jessen, Philipp Ströbel, Bernd Auber, Felix Sahm, Sabine Merkelbach-Bruse, Udo Siebolts, Wilfried Roth, Silke Lassmann, Frederick Klauschen, Nadine T. Gaisa, Wilko Weichert, Matthias Evert, Sorin Armeanu-Ebinger, Stephan Ossowski, Christopher Schroeder, Christian P. Schaaf, Nisar Malek, Peter Schirmacher, Daniel Kazdal, Nicole Pfarr, Jan Budczies and Albrecht Stenzinger, Benchmarking Whole Exome Sequencing in the German Network for Personalized Medicine, *European Journal of Cancer,* (2024) doi:https://doi.org/10.1016/j.ejca.2024.114306

Benchmarking whole exome sequencing in the German Network for Personalized Medicine

Michael Menzel[1,2], Mihaela Martis-Thiele[3], Hannah Goldschmid[1,2], Alexander Ott[4], Eva Romanovsky[1,2], Janna Siemanowski-Hrach[5], Lancelot Seillier[6,7,8], Nadina Ortiz Brüchle[6,8], Angela Maurer[6], Kjong-Van Lehmann[7,8,9,10], Matthias Begemann[11,12,8], Miriam Elbracht[11,8], Robert Meyer[13,14,8], Sebastian Dintner[15], Rainer Claus[15,16], Jan Meier-Kolthoff[17], Eric Blanc[18], Markus Möbs[19], Maria Joosten[19], Manuela Benary[18,20], Patrick Basitta[21], Florian Hölscher[21], Verena Tischler[21], Thomas Groß[22], Oliver Kutz[23], Rebecca Prause[22], Doreen William[22,23], Kai Horny[24,25], Wolfgang Goering[26], Sugirthan Sivalingam[27], Arndt Borkhardt[28], Cornelia Blank[27], Stefanie Junk[28], Layal Yasin[28], Evgeny A. Moskalev[29,30,31,32], Maria Giulia Carta[29,30,31,32], Fulvia Ferrazzi[29,30,31,32,33], Lars Tögel[29,30,31,32], Steffen Wolter[34,35,36], Eugen Adam[34,35,36], Uta Matysiak[34,35,36], Tessa Rosenthal[37], Jürgen Dönitz[38], Ulrich Lehmann[39], Gunnar Schmidt[40], Stephan Bartels[39], Winfried Hofmann[41], Steffen Hirsch[42], Nicola Dikow[42], Kirsten Göbel[43], Rouzbeh Banan[43], Stefan Hamelmann[43], Annette Fink[1,2], Markus Ball[1,44], Olaf Neumann[1,2], Jan Rehker[5], Michael Kloth[45], Justin Murtagh[45], Nils Hartmann[45], Phillip Jurmeister[46,47], Andreas Mock[46,47], Jörg Kumbrink[46,47], Andreas Jung[46,47], Eva-Maria Mayr[3], Anne Jacob[3], Marcel Trautmann[48,49], Santina Kirmse[48,49], Kim Falkenberg[48,49], Christian Ruckert[50], Daniela Hirsch[51], Alexander Immel[52,51], Wolfgang Dietmaier[51], Tobias Haack[4], Ralf Marienfeld[53,54], Axel Fürstberger[53,54], Jakob Niewöhner[53,54], Uwe Gerstenmaier[53,54], Timo Eberhardt[53,54], Phillip Greif[55,56,47], Silke Appenzeller[57], Katja Maurus[58], Julia Doll[58], Yvonne Jelting[59], Danny Jonigk[6,8,60], Bruno Märkl[15], Dieter Beule[18], David Horst[19], Anna-Lena Wulf[21], Daniela Aust[22,61], Martin Werner[34,35,36,62], Kirsten Reuter-Jessen[37], Philipp Ströbel[37], Bernd Auber[40], Felix Sahm[43], Sabine Merkelbach-Bruse[5], Udo Siebolts[5], Wilfried Roth[45], Silke Lassmann[63,64], Frederick Klauschen[46,47], Nadine T. Gaisa[53,6,8], Wilko Weichert[3,+], Matthias Evert[51], Sorin Armeanu-Ebinger[4], Stephan Ossowski[4], Christopher Schroeder[4], Christian P. Schaaf[42], Nisar Malek[2,65], Peter Schirmacher[1,2], Daniel Kazdal*[1,44], Nicole Pfarr*[3], Jan Budczies*[1,2], Albrecht Stenzinger*[1,2,44,66]

* Shared last author [+]deceased

[1]Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany, [2]Centers for Personalized Medicine (ZPM), Germany, [3]Institute of Pathology, TUM School of Medicine and Health, Technical University of Munich, [4]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany, [5]Institute of Pathology, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany, [6]Institute of Pathology, University Hospital RWTH Aachen, Aachen, Germany, [7]Joint Research Center Computational Biomedicine, University Hospital RWTH Aachen, Aachen, Germany, [8]Center for Integrated Oncology Aachen Bonn Cologne Düsseldorf (CIO ABCD), [9]Cancer Research Center Cologne-Essen, University Hospital Cologne, [10]Machine Learning in Cancer Genetis and Precision Medicine, University RWTH Aachen, Aachen, Germany, [11]Institute for Human Genetics and Genomic Medicine., University Hospital RWTH Aachen, Aachen, Germany, [12]NGS diagnostic centre, Uniklinik RWTH Aachen, Aachen, Germany, [13]Institut für Humangenetik und Genommedizin, Uniklinik RWTH Aachen, Aachen, Germany, [14]NGS-Diagnostikzentrum, Uniklinik RWTH Aachen, Aachen, Germany, [15]Pathology, Medical Faculty Augsburg, University Augsburg, Germany, [16]Comprehensive Cancer Center, Medical Faculty Augsburg, University Augsburg, Germany, [17]Chair of Biomedical Informatics, Data Mining and Data Analytics, Faculty of Applied Computer Science, University of Augsburg, Germany, [18]Core Unit Bioinformatics, Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Charitéplatz 1, Berlin, Germany, [19]Institute of Pathology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany, [20]Charité Comprehensive Cancer Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany, [21]Universitätsklinikum Bonn, Molekularpathologische Diagnostik, Institut für Pathologie, Venusberg Campus 1, 53127 Bonn, [22]Core Unit for Molecular Tumor Diagnostics (CMTD), National Center for Tumor Diseases Dresden (NCT), NCT/UCC Dresden, a partnership between German Cancer Research Center (DKFZ), Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology and Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany, [23]Institute for Clinical Genetics, University Hospital Carl Gustav Carus at TUD Dresden University of Technology and Faculty of Medicine of TUD Dresden University of Technology, Dresden, Germany; ERN GENTURIS, Hereditary Cancer Syndrome Center Dresden, Germany; National Center for Tumor Diseases (NCT), NCT/UCC Dresden, a partnership between German Cancer Research Center (DKFZ), Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology and Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany; German Cancer Consortium (DKTK), Dresden, Germany; German Cancer Research Center (DKFZ), Heidelberg, Germany; Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, [24]Center for Personalized Medicine Oncology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Germany, [25]Core Unit Bioinformatics, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Germany, [26]Institute of Pathology, University Hospital Duesseldorf, [27]Institute of Human Genetics, Medical Faculty, University Hospital of Düsseldorf, Heinrich Heine University of Düsseldorf, Düsseldorf, Germany, [28]Department of Pediatric Oncology, Haematology and Clinical Immunology, Medical Faculty, HHU Dusseldorf, [29]Institute of Pathology, Friedrich-Alexander-Universität Erlangen-Nürnberg, [30]Centers for Personalized Medicine (ZPM), Erlangen, Germany, [31]Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen, Germany, [32]Bavarian Cancer Research Center (BZKF), Erlangen, Germany, [33]Department of Nephropathology, Institute of Pathology, Friedrich-Alexander-Universität Erlangen-Nürnberg, [34]Institute for Surgical Pathology, Medical Center, University of Freiburg, Germany, [35]Center for Personalized Medicine (ZPM), partner site Freiburg, Germany, [36]Comprehensive Cancer Center Freiburg (CCCF), Medical Center, Freiburg, Germany, [37]Institut für Pathologie, Universitätsmedizin Göttingen, [38]Institut für Bioinformatik, Universitätsmedizin Göttingen, [39]Institute of Pathology, Hannover Medical School, Hannover, Germany, [40]Institute of Human Genetics, Hannover Medical School, Hannover, Germany, [41]Department of

Human Genetics, Hannover Medical School, Hannover, Germany, [42]Institute of Human Genetics, Heidelberg University, Heidelberg, Germany, [43]Department of Neuropathology, University Hospital Heidelberg, Germany, [44]Translational Lung Research Center Heidelberg (TLRC), Member of the German Center for Lung Research (DZL), Heidelberg, Germany, [45]Institut für Pathologie, Universitätsmedizin Mainz, Germany, [46]Institute of Pathology, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany, [47]German Cancer Consortium, German Cancer Research Center (DKTK/DKFZ), Munich, Partner Site, Munich, Germany, [48]Gerhard-Domagk-Institute of Pathology, University Hospital Münster, Münster, Germany, [49]West German Cancer Center, University Hospital Münster, Münster, Germany, [50]Centre of Medical Genetics, Department of Medical Genetics, University and University Hospital Münster, Münster, Germany, [51]Institute of Pathology, University of Regensburg, [52]Centrum für Translationale Onkologie, Universitätsklinikum Regensburg, [53]Institute of Pathology, University Hospital Ulm, [54]Centers for Personalized Medicine (ZPM), Ulm, Germany, [55]Department of Medicine III, University Hospital, LMU Munich, Munich, Germany, [56]Institute of Human Genetics, University Hospital, LMU Munich, Munich, Germany, [57]Comprehensive Cancer Center Mainfranken, University Hospital Wuerzburg, [58]Institute of Pathology, University of Wuerzburg, [59]Institute of Human Genetics, University of Wuerzburg, [60]Biomedical Research in End-stage and Obstructive Lung Disease Hannover (BREATH), German Lung Research Centre (DZL), Hannover, Germany, [61]Institut für Pathologie, Universitätsklinikum Carl Gustav Carus der TU Dresden, Fetscherstr. 74, 01307 Dresden., [62]German Cancer Consortium (DKTK), partner site Freiburg, Germany, [63]Institute for Surgical Pathology, Medical Center, University of Freiburg, Freiburg, Germany, [64]Center for Personalized Medicine (ZPM), Freiburg, Germany, [65]Department of Gastroenterology, Tübingen University Hospital, Tübingen, Germany, [66]German Cancer Consortium (DKTK), Germany

Correspondence to:

Dr. rer. nat. Michael Menzel, Institute of Pathology, Heidelberg University Hospital, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany. Tel.: +49.6221.310536, Email: michael.menzel@med.uni-heidelberg.de

and

Prof. Dr. med. Albrecht Stenzinger, Institute of Pathology, Heidelberg University Hospital, Im Neuenheimer Feld 224, 69120 Heidelberg, Germany. Tel.: +49.6221.5634380, Email: albrecht.stenzinger@med.uni-heidelberg.de

Abstract

Introduction: Whole Exome Sequencing (WES) has emerged as an efficient tool in clinical cancer diagnostics to broaden the scope from panel-based diagnostics to screening of all genes and enabling robust determination of complex biomarkers in a single analysis.

Methods: To assess concordance, six formalin-fixed paraffin-embedded (FFPE) tissue specimens and four commercial reference standards were analyzed by WES as matched tumor-normal DNA at 21 NGS centers in Germany, each employing local wet-lab and bioinformatics investigating somatic

and germline variants, copy-number alteration (CNA), and different complex biomarkers. Somatic variant calling was performed in 494 diagnostically relevant cancer genes. In addition, all raw data were re-analyzed with a central bioinformatic pipeline to separate wet- and dry-lab variability.

Results: The mean positive percentage agreement (PPA) of somatic variant calling was 76% and positive predictive value (PPV) 89% compared a consensus list of variants found by at least five centers. Variant filtering was identified as the main cause for divergent variant calls. Adjusting filter criteria and re-analysis increased the PPA to 88% for all and 97% for clinically relevant variants. CNA calls were concordant for 82% of genomic regions. Calls of homologous recombination deficiency (HRD), tumor mutational burden (TMB), and microsatellite instability (MSI) status were concordant for 94%, 93%, and 93% respectively. Variability of CNAs and complex biomarkers did not increase considerably using the central pipeline and was hence attributed to wet-lab differences.

Conclusion: Continuous optimization of bioinformatic workflows and participating in round robin tests are recommend.

Keywords: whole exome sequencing; molecular pathology; multi-centric inter-laboratory test; clinical exome; precision oncology

Introduction

Currently, the implementation of clinical Whole Exome Sequencing (WES) in predictive molecular cancer diagnostics is expedited by decreasing sequencing costs and government reimbursement schemes in Germany. WES offers several substantial advantages[1] over sequencing panels, such as eliminating the risk to miss targetable alterations in the coding sequences of the genome and

enabling *post-hoc* retrospective research in an unrestricted manner once new putative targets emerge. Further, using WES provides a more robust measurement of complex biomarkers[2,3].

To evaluate reproducibility between different pipelines and uncover opportunities for improvement of workflows, we compared an unprecedented number of diagnostic centers regarding their initial implementation of WES analysis. Previous WES implementation studies with clinical focus were either single center[4–6], or focused on limited metrics, such as single complex biomarkers or only variant calling[7–17]. An earlier study performed in Germany included a smaller number of laboratories investigated WES of fresh frozen tissue samples[1]. WES for cancer patients from FFPE samples has not been compared between a large number of centers before.

Six clinical FFPE tissue specimens and four commercial reference samples of matched tumor and normal DNA samples were analyzed in the German National Initiative for Personalized Medicine (DNPM) at 21 participating centers, using locally established wet-lab workflows and dry-lab bioinformatics pipelines. Somatic and germline mutations in diagnostically relevant genes, complex biomarkers, and copy-number alterations (CNAs) were reported and evaluated for concordance. Discrepancies were cooperatively assessed to identify relevant factors for optimization and harmonization of WES in clinical cancer diagnostics. Additionally, raw sequencing data sets were re-analyzed using a central bioinformatic pipeline to separate wet- and dry-lab variability. This study presents the results of a national comprehensive FFPE tissue specimen based evaluation for WES analysis in a clinical routine diagnostic setting.

Material and Methods

Six FFPE tissue specimens were selected to represent various cancer types and biological features (Table 1). Broad consent was given by the patients and all analysis were performed in line with the Declaration of Helsinki. Further, four reference samples were included with two cases of low and high tumor mutational burden (TMB), as well as two cases of low and high Homologous Recombination Deficiency (HRD) score. Matched tumor-normal DNA from FFPE tissue specimens was extracted centrally. DNA from the FFPE specimen and the reference samples were shipped to the 21 participants for wet- and dry-lab analysis.

Each center used their wet and dry-lab protocols to analyze a defined set of parameters, including somatic and germline variants, CNA, and complex biomarkers. Next, the results were gathered and compared for concordance and against the known values of the reference samples. Raw data was collected as unfiltered variant calls and unaligned reads. Those were re-analyzed using the pipeline of Center-5 to separate effects of wet and dry-lab. Detailed information can be found in the Supplementary Methods.

Results

WES was performed using local wet-lab and bioinformatic pipelines (Supplementary Table 1). Most centers successfully sequenced all of the samples, while Center-1 and Center-11 excluded samples 3 and 16 because of sequencing quality issues.

Across 18 centers and the sequenced tumor samples, the median mean target coverage was 268x (lower quartile 218, upper quartile 354). A total of 154 (88%) samples were covered with >100x for at least 80% of the target region (Supplementary Figure 1). Of the remaining three centers, two (Center-10 and Center-12) had a lower mean coverage (87x and 73x), while Center-19 had a considerably higher coverage (1374x).

Somatic variants

Somatic variants in the predefined list of 454 diagnostically relevant cancer genes (Supplementary Table 2) were compared by chromosomal position and base alteration. Based on a total of 1014 unique non-synonymous somatic variants detected by 5,799 variant calls, we defined a consensus list of 321 variants that were called by at least five centers (Figure 1a). The distribution of missed variant calls shows that more variant calls (60%) where missed by few (1-5 centers), while 40% of calls where missed by more centers (Supplementary Figure 2a). Across centers, a positive percentage agreement of 76% (5,000 variant calls) was reached with respect to this list. By contrast, a total of 1,590 (24%) variant calls were missed by one or more centers which were examined in an in-depth analysis to uncover the causes for the discrepancies (Figure 1a). Of the missed variants, 750 (47%) could be found by reviewing the unfiltered variant calls and were reported to the respective centers for review. Variant filters, including PASS filter (18.3%) and low variant allele fraction (24.4%) were reported as predominant causes for not reporting (Figure 1b). Furthermore, we uncovered several local filter rules that were only used in a single or in few centers (Supplementary Figure 2b). Finally, several missed calls were found to be due to incorrect filter rules and reported as erroneous.

For 31% of the reported variant calls the VAF was below 5% VAF. In the consensus list only four variants (1%) were below this threshold as many centers filtered with a fixed VAF threshold. We evaluated the influence of a central VAF cut-off on PPV and PPA. For a 5% VAF cut-off the mean PPV increased from 89% to 91%, due to the removal of possible false-positives, the mean PPA increased from 74% to 75%. For a 10% VAF cut-off the mean PPV increased further to 95%, while the mean PPA increased from 74% to 80% as more consensus variants were found between 5-10% VAF (55 Variants, 17%).

Of the remaining 829 missed calls, 580 (70%) could be identified by the re-analysis of raw-data with the central bioinformatics pipeline. An additional 127 variants were found but filtered out by either depth, variant allele fraction (VAF) or PASS filter. Overall, only 122 (1.9%) of all expected somatic variant calls were not identified by either initial analysis, raw variant calls or re-analysis. Those variants showed a low mean VAF of 2.4%. The fractions of found to miss ratio varied among the centers (Figure 1c). The set of missed variant calls showed a significant enrichment of InDels (Supplementary Figure 3a). The missed InDels were often located at homopolymer sites (120 of 144 = 83%). Further, a negative correlation of PPA with the sample sequencing depth was observed (slope=-0.011, p=0.007, Supplementary Figure 3b).

A total of 628 variants were only reported by single centers (Figure 1a) of which all but 21 variants could be classified as either low VAF (< 5% or < 10%), low depth (< 100 reads), InDels or variants at homopolymer sites. Most of these variants were reported by centers 11, 13, 16, 20 and especially 19, which reported the most variants which was in line with a much higher sequencing coverage.

In the absence of a gold-standard for somatic variants in the analyzed cases different consensus lists were selected based on the number of missed calls. For each of the lists, PPA and positive predictive value (PPV) were determined. The median PPA for the five-center consensus list (16 missed calls allowed) was 75%, while a median PPV of 89% was reached, (Figure 2a,b). Considerably lower PPVs were observed for the five centers that submitted the highest numbers of variants (listed above).

Based on the five-center consensus list, the unfiltered variant calls of each center were searched for the missed calls. Including these variants lead to a considerable increase of PPA for previously lower performing centers (Figure 2c, middle). Re-analysis of raw data using the same bioinformatics pipeline further increased the PPA to a mean of 88%. This correction especially improved the worst performing centers (Figure 2c, right). Clustering of centers by detected somatic variants showed no relation to wet- or dry-lab procedures (Supplementary Figure 4). No correlation between self-reported experience with WES analysis and variant calling performance (PPA, PPV) was observed (p values between 0.3 - 0.6).

The five center consensus list was screened for druggable targets using OncoKB without considering cancer type[18]. Altogether, there were 31 druggable variants, of which on average 80% were identified by the centers (523 calls). An additional 17% (111 calls) could be found either in the list of raw variant calls or with the re-analysis of raw-data as described previously, resulting in total of 97% of the druggable variants that were identified (Figure 3).

Germline variants

In total, 10 centers reported pathogenic and likely pathogenic germline variants and their classification. The consensus results, i.e., variants identified by at least five centers, included two (likely) pathogenic variants in *PMS2* (p.E504X) and *BRCA1* (p.Q1756Pfs*74) and two variants classified as (likely) benign or of unknown significance (VUS) in *RET* (p.Y791F) and *TP53* (p.R283C), reported as likely pathogenic by one center (Supplementary Figure 5a). The *BRCA1* variant was identified by all centers. The *PMS2* variant was not reported by two centers due to a pseudogene filter. The variants in *RET* and *TP53* were each identified by all but one center.

All of the centers classified the germline variants according to ACMG criteria. Classification of the *BRCA1* variant was concordant and reported as pathogenic by nine centers and as likely pathogenic by the remaining center. The classification of the *PMS2* variant was more heterogeneous, with a consensus classification as likely pathogenic (pathogenic: 2 centers, likely pathogenic: 6 centers, quality filtered: 2 centers). The consensus classification for the variants in *RET* and *TP53* was benign (7 out of 10 centers in each case), while being reported as either VUS or likely pathogenic by two centers. ACMG criteria for the pathogenic variants in *BRCA1* and *PMS2* showed high similarity (Supplementary Figure 5b), with differences leading to differential classification of the *PMS2* variant as pathogenic or likely pathogenic. In line with the detected (likely) pathogenic variants in *BRCA1* and *PMS2*, high HRD and MSI scores were observed for the corresponding tumors, respectively.

Somatic copy number alterations

Genome-wide allele-specific CNA segments with absolute CN were submitted by 18 centers. Pairwise comparison of genomic regions across the whole genome lead to an agreement of 61% of bases with 11% of bases matching when accounting for genome duplications, 28% of bases showed divergent values (Figure 4a). Re-analysis of raw data with a single pipeline improved the concordance to 72% match and while duplication match decreased to 7% (Supplementary Figure 6a). Hierarchical clustering of CNA revealed three main clusters, which can be attributed to differences by bioinformatics tools (Figure 4b), re-analysis of raw data with a central pipeline led to a clustering by panel for most centers (Supplementary Figure 6b).

Concordance of gene amplifications, deep deletion and LOH calls was calculated in reference to alterations found by at least five centers in genes of the somatic gene list. While the mean PPA for the detection of amplifications was moderate (59%), a mean PPV of 77% was achieved. Concordance of deep and LOH calls were observed to be higher with a mean PPV of 81% and mean PPV of 82% (Figure 4d). Re-analysis of raw data with a single pipeline improved the PPA for detection of amplifications only by 10%, while the PPV did not change. Similarly, PPA for the detection of deletions was improved by only 1%, PPV remained unchanged (Supplementary Figure 6c,d).

Gene specific copy number alterations were investigated for genes with level 1-4 of OncoKb[18] and found two elevated CN for *MDM2* and *MET*, as well as two high-level amplifications: sample 3 *FGFR1* with a median of 17 copies found by 16 of 18 centers and *ERBB2* in sample 20 with a

median of 19 copies found by 16 of 18 centers (Supplementary Figure 7a). A deep deletion of *CDKN2A* was found by 6 centers in sample 3 as well as a varying counts of losses for *CDKN2A* in sample 17 and *TP53* in sample 20 (Supplementary Figure 7b).

Complex biomarkers

HRD scores were determined using eight different bioinformatics segmentation tools (Figure 5a) and three different methods to count genomic scars. Fourteen centers reported results using the commonly used cut-off of 42 inferred from breast and ovarian carcinoma[19], one center used a cut-off of 65, while the six remaining centers did not perform HRD classification (Supplementary Table 1). Unanimous status calls were observed for four of 10 samples. Overall, 134 (93%) of the status calls were consistent across centers. Identical status calls were reported for the reference samples (Figure 5a). Correlations of HRD scores between centers showed a mean of 0.88 ± 0.18 with a 75% percentile above 0.98, with only Center-1 showing correlations below 0.69, which applied a different cut-off and used a different bioinformatics tool (Supplementary Figure 8).

Overall, 163 TMB status calls (93%) showed agreement, based on the cut-off 10 Mut/Mb for TMB-high vs. low (Figure 5b). Most centers were in agreement of the status calls, but some center-specific discordance were observed: Center-11 showed overall lower TMB values, this center also had the most missed somatic variant calls. Center-20 showed considerably higher TMB values. Other deviations from the consensus were close to the cut-off point. For the two reference samples, 33 status calls (86%) were concordant over all centers. Correlation of TMB values showed a mean of 0.89 ± 0.2 with 99% of calls in the 75% percentile.

Four different bioinformatics tools were utilized for the calculation of MSI scores in 20 centers, one center did not submit MSI values. MSI status was determined with different cut-off values. Six centers did not submit a cut-off value. In seven samples full concordance was observed, while overall 132 (94%) status calls were concordant (Figure 5c). The majority status calls were consistent with assay-based (qPCR or gold standard fragment length analysis) and TSO500 status calls. MSI values showed the highest mean correlation (0.96 ± 0.07) with only Center-9 displaying a correlation coefficient below 0.9.

Re-analysis of biomarkers from the raw data with the central pipeline revealed an improvement of status calls concordance for MSI (+2%pt), while no improvement for TMB (+0%pt) and decrease for HRD status calls (-3%pt) were observed (Supplementary Figure 9). Comparative Interclass correlation (ICC) between original and central bioinformatics showed varying results for the original data (TMB: 0.30, MSI: 0.59, HRD: 0.77). The largest impact on variance was found to be center-specific as TMB and MSI were improved distinctly by removing two highly variant centers in TMB (Center-11 and Center-20, increased ICC from 0.30 to 0.77) and five centers with different cut-offs in MSI (increased from 0.59 to 0.72) (Figure 5d). Changes in ICC between original and central bioinformatics were comparable to the improvements of status calls discussed before with better concordance for MSI (0.59/0.72 to 0.92) and slight changes in TMB (0.30/0.77 to 0.77) and HRD (0.77 to 0.77) (Figure 5d).

A total of 67 different single base substitution (SBS) signatures were submitted by 10 centers using three different bioinformatics tools showing a mean ICC of 0.36 (Figure 5e, original). Re-analysis of all raw data with two bioinformatics tools substantially improved the mean ICC to 0.73 and 0.88

(Figure 5e, central). ICC comparing both central methods shows high correlation for the signatures SBS2, SBS4 and SBS13. Lower ICC was observed for the more prevalent signatures SBS1 and SBS5. The signature SBS6 was called, yet, in different samples (Figure 5e, central).

HLA class I status was determined by 10 centers using six different bioinformatic tools. Overall, 576 of 660 (87%) HLA calls were concordant. Hierarchical clustering indicated a strong correlation of HLA predictions for all but one tool (Supplementary Figure 10).

Discussion

As data on inter-center comparability of diagnostic WES using FFPE material are scarce[1], we initiated a national benchmark study involving 21 major cancer centers in Germany. The DNA of six paired tumor and normal specimen and four reference samples was analyzed with local wet-lab workflows and bioinformatics pipelines, as well as a central bioinformatics pipeline to allow separation between wet-lab and dry-lab variability.

Somatic variant calls showed an average PPA of 76% compared to the five-center consensus list. Deviations could largely be explained by different variant filter rules. Re-analysis revealed that in principle an average PPA of 98% was achievable from the raw-data. Therapeutically relevant variants reached an average PPA of 80%, which potentially could be improved to 97%. An influence of FFPE on concordances was not observed[1].

Based on unfiltered somatic variant calls and re-analysis we were able to determine the main factors for the four centers with overall lower concordance: Center-1 used different bioinformatics tools (Supplementary Table 1) and most variant calls could be found in the re-analysis (Figure 1c).

Center-10 missed most variant calls due to lower coverage (87x) as variants were found, but often not labeled as PASS (Supplementary Figure 2a). Center-3 implied a strict manual filter accountable for about two-thirds of missed calls while most others were filtered due to a misconfigured pipeline (Supplementary Figure 2b). Center-20 implied strict variant filtering and variants could often be found in the unfiltered calls (Figure 2c).

Four germline variants in cancer risk genes were identified and classified with high concordance between centers, with some notable exceptions. Two centers did not report the variant in *PMS2*, which lies in a homologous region on exon 11 (repetitive due to pseudogene PMS2CL) and was therefore removed by the center's quality filters. While most centers agreed that the variants in *RET* and *TP53* are (likely) benign, one center classified the variants as likely pathogenic. At the time of the round robin test the two variants were annotated as likely pathogenic in at least one of the three databases OnkoKB, CKB and LOVD used by the center and reported with the highest observed class. In the meantime, the *RET* variant class was reduced to 2-3 in all three databases, while the *TP53* variant remained a class 4 only in CKB, showing that pathogenicity classification is highly dependent on utilized databases.

CNA concordance was observed for 72% of the genomic regions, which was very similar to 76% of matching regions observed in the earlier study in fresh frozen tissue specimens[1].
Deviations between bioinformatic tools were also observed, a results that ties well with a systematic evaluation of wet-lab influences and bioinformatics evaluation on CNV calling[20]. The differences between centers were not resolved by using the same bioinformatics tools as seen for SNVs. Therefore, the underlying cause appears to be wet-lab driven. In line with this notion, a

recent study showed that WES and FFPE processing had a large impact on CNV concordance, especially on losses[20].

Status calls for HRD, TMB, and MSI agreed for 93%, 93%, and 94% across samples and centers, respectively, even though different bioinformatics tools were used. The results align with previous results in fresh frozen tissue specimens (HRD: 96%, TMB: 99%, MSI: 100%). Re-analysis in the central pipeline did not increase the concordances beyond well-aligned wet- and dry-lab procedures. The estimation of HRD scores is strongly influenced by the estimation of tumor purity and ploidy[21]. For the artificial reference sample 17 different ploidy solutions between 1.8 and 4.4 were chosen, which presumably lead to deviant results of the HRD score estimation while results were around the cut-off value of 42 further increasing the discrepancies. The re-analysis showed a higher concordance but the values are also scatter around the cut-off (Supplementary Figure 9). The influence of the segmentation tool, as shown in Figure 2a, seems to be minor, except for Center-1, as neither a systematic deviation is apparent between the tools, nor did the re-analysis improve HRD scores. Re-analysis of MSI still showed large deviations in status for sample 14, where also a high variance in bioinformatically estimated tumor purity between 10-100% were observed, emphasizing the difficulties in evaluating this sample (Supplementary Figure 9).

Aside from the center-specific deviations described before, increasing sequencing depth correlated with less missed variant calls (Supplementary Figure 3). However, no clustering by sequencing depth was observed (Supplementary Figure 4), indicating that low sequencing depth lead to less variant identification, yet, beyond a certain depth, it does not increase concordance. No significant correlation between sequencing depth and PPV or PPA of CNA were found (p values between 0.2 - 0.6).

Findings in this study are limited by the absence of a gold standard for somatic mutations, complex biomarkers or CNA for most cases. Missed somatic variants could be recovered in unfiltered VCFs, however, this approach needs to be balanced with false positive detection. Furthermore, it should be noted that the central bioinformatics pipeline used in the re-analysis of raw data only represents a single possible approach. Other bioinformatics tools could improve the concordance further.

Previous studies have laid the foundation for clinical WES[22] and highlighted the benefits of moving from gene panels to WES, which allows for rapid and flexible expansion of the reportable gene list and precise measurement of complex biomarkers while reducing the burden of assay revalidation[23]. Our multicentric benchmark study, which is to our best knowledge the largest of its kind, closes a significant gap in the field, supports the implementation of decentralized WES in clinical diagnostics for cancer patients and demonstrates its fundamental feasibility. The results also highlight processes in the dry laboratory that require further standardization and harmonization. Furthermore, our study also provides a basis and blueprint for the design of standardized EQA schemes for clinical WES.

Data Availability

The datasets generated during the current study are available from the corresponding authors on reasonable request.

References

1. Menzel M, Ossowski S, Kral S, et al. Multicentric pilot study to standardize clinical whole exome sequencing (WES) for cancer patients. *NPJ Precis Oncol*. 2023;7(1):106. doi:10.1038/s41698-023-00457-x

2. Rempel E, Kluck K, Beck S, et al. Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (HRD). *Npj Precis Oncol*. 2022;6(1):1-13. doi:10.1038/s41698-022-00276-6

3. Budczies, J et al. Optimizing panel-based tumor mutational burden (TMB) measurement. *Annals of oncology : official journal of the European Society for Medical Oncology* vol. 30,9 (2019): 1496-1506. doi:10.1093/annonc/mdz205

4. Ramarao-Milne P, Kondrashova O, Patch A-M, et al. Comparison of actionable events detected in cancer genomes by whole-genome sequencing, in silico whole-exome and mutation panels. *ESMO Open*. 2022;7(4):100540. doi:10.1016/j.esmoop.2022.100540

5. Auzanneau C, Bacq D, Bellera 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*. 2020;5(4). doi:10.1136/esmoopen-2020-000744

6. Massard C, Michiels S, Ferté C, et al. High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial. *Cancer Discov*. 2017;7(6):586-595. doi:10.1158/2159-8290.CD-16-1396

7. 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*. 2016;6(1):36540. doi:10.1038/srep36540

8. Krøigård AB, Thomassen M, Lænkholm A-V, Kruse TA, Larsen MJ. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PLOS ONE*. 2016;11(3):1-15. doi:10.1371/journal.pone.0151664

9. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 1998;8(3):186-194.

10. Zhao Y, Fang LT, Shen T-W, et al. Whole genome and exome sequencing reference datasets from a multi-center and cross-platform benchmark study. *Sci Data*. 2021;8(1):296. doi:10.1038/s41597-021-01077-5

11. Xiao W, Ren L, Chen Z, et al. Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing. *Nat Biotechnol*. 2021;39(9):1141-1150. doi:10.1038/s41587-021-00994-5

12. Barbitoff YA, Abasov R, Tvorogova VE, Glotov AS, Predeus AV. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. *BMC Genomics*. 2022;23(1):155. doi:10.1186/s12864-022-08365-3

13. Gabrielaite M, Torp MH, Rasmussen MS, et al. A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. *Cancers*. 2021;13(24). doi:10.3390/cancers13246283

14. Merino DM, McShane LM, Fabrizio D, 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*. 2020;8(1):e000147. doi:10.1136/jitc-2019-000147

15. Vega DM, Yee LM, McShane LM, et al. Aligning tumor mutational burden (TMB) quantification across diagnostic platforms: phase II of the Friends of Cancer Research TMB Harmonization Project. *Ann Oncol Off J Eur Soc Med Oncol*. 2021;32(12):1626-1636. doi:10.1016/j.annonc.2021.09.016

16. Lambin S, Lambrechts D, De Rop C, 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*. 2019;30:xi10. doi:10.1093/annonc/mdz447.031

17. Velasco A, Tokat F, Bonde J, 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*. 2021;478(5):851-863. doi:10.1007/s00428-020-02962-x

18. Chakravarty D, Gao J, Phillips S, et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol*. 2017;(1):1-16. doi:10.1200/PO.17.00011

19. Telli ML, Timms KM, Reid J, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2016;22(15):3764-3773. doi:10.1158/1078-0432.CCR-15-2477

20. Masood, D., Ren, L., Nguyen, C. *et al.* Evaluation of somatic copy number variation detection by NGS technologies and bioinformatics tools on a hyper-diploid cancer genome. *Genome Biol* **25**, 163 (2024). https://doi.org/10.1186/s13059-024-03294-8

21. Menzel M, Endris V, Schwab C, et al. Accurate tumor purity determination is critical for the analysis of homologous recombination deficiency (HRD). *Translational oncology* vol. 35 (2023): 101706. doi:10.1016/j.tranon.2023.101706

22. Van Allen EM, Wagle N, Stojanov P, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med*. 2014;20(6):682-688. doi:10.1038/nm.3559

23. Shah PS, Hughes EG, Sukhadia SS, et al. Validation and Implementation of a Somatic-Only Tumor Exome for Routine Clinical Application. *J Mol Diagn*. Published online July 6, 2024. doi:10.1016/j.jmoldx.2024.05.013

Acknowledgments

We thank Twist Bioscience, Qiagen, IDT, Illumina, and Agilent for the partial provisioning of sequencing reagents and Seracare for the partial provisioning of reference material and GenXPro for individual bioinformatic service. We thank the QuIP (Ms. Ilm) for sample handling and logistics.

Author contributions

MM, MMT, HGO, DK, NP, JB, and AS planned and supervised the study. MM, DK, NP, JB, and AS drafted the manuscript. MM, MMT, HGO, ET, JS, ON, MB, AFI, JR, EM, AJA, SMB, WW, PS, DK, NP, JB, AS provided samples. MM, MMT, HGO, A.O., ET, JS, LS, NOB, AM, KVL, MB, ME, RM, SD, RC, JMK, EB, MMO, MJ, MBE, PB, FH, VT, TG, OK, RP, DW, KH, WG, SuS, AB, CB, SJ, LY, EAM, MGC, FF, LT, S.W., E.A., U.M., TR, JUD, UL, GS, SB, WH, SH, ND, KG, RB, SHA, AFI, MB, ON, JR, MK, JM, NH, PJ, AM, JK, AJ, EM, AJA, MT, SK, KF, CR, DH, AI, WD, T.B.H., RM, AF, JN, UG, TE, PG, SA, KM, JD, YJ, DJ, BM, DB, DHO, ALW, DA, M.W., KRJ, PST, BA, FS, SMB, Usi, WR, SL, FK, NTG, WW, ME, S.A-E., S.O., CS, CPS, NM, PS, DK, NP, JB, and AS contributed and analyzed data, contributed to the writing and approved the final manuscript.

Funding

This study was partly supported by the DNPM.

Figure captions

Figure 1: Somatic variant calls. a: Variants separated by occurrence. b: Reasons reported by centers for missed variants. c: Fractions of missed variants found by another method or not found separated by center.

Figure 2: Positive percentage agreement (PPA) and positive predictive values (PPV) of variant calls. a: PPA in relation to allowed missed calls and inversely to the consensus counts. For zero missed calls a consensus of all 21 centers is found, for 16 allowed misses a consensus of five centers is found. The consensus counts were created by selecting the variants missed by at most the number of centers annotated at the bottom. The top shows the number of variants for each of the consensus counts. b: PPV in relation to decreasing consensus counts. c: Change in PPA for unfiltered calls (middle), or the single re-analysis with basic filters (PASS filter, VAF >= 3.5%, depth >= 100) in relation to the five center consensus.

Figure 3: Variants filtered for possible therapeutic targets. Each box shows a reported calls, they are colored for better samples separation. 51 missed variant calls were found in the raw calls and were annotated with the reason for the miss as reported by the respective center. Further 60

variants were found using the central bioinformatic pipeline. White fields indicate that neither local nor central could find the variant call.

Figure 4: Comparison of CNA calls. a: Pairwise comparison of CN profiles by sample and separated into segments with matching (green), not matching (red), and matching when normalized for genome duplications (purple). b: Hierarchical clustering of CN profiles annotated with bioinformatic segmentation tool. c: Gene amplification calls by center in relation to the five center consensus. d: Gene deep deletion or LOH calls in relation to the five center consensus.

Figure 5: Reported values for complex biomarkers. a: HRD scores ordered by segmentation tool, cut-off 42 was used in all but Center-1. b: Missense TMB values colored by TMB status with 10 Mut/MB cut-off. c: MSI percentage unstable sites ordered by bioinformatic tool with the cut-off value. The rightmost column shows the fraction of instable sites by MSI assay. d: Interclass correlation (ICC) for the three biomarkers from both bioinformatic pipelines. e: ICC for the main mutational signatures from the original data, the two algorithms from the central pipeline, and between the two algorithms.

Tables

Table 1: Samples used in the pilot study with previously determined characteristics from TSO500 (samples 3 - 16) and reference material (samples 17 – 20). Complex biomarker were were not determined (N.D.) for all samples. *CytoSNP; $^{+}$confirmed by Marker panel; ‡determined by Seraseq.

Sample ID	Entity / Reference	Pathological tumor purity (%)	HRD	TMB	MSI (%)
3	Undifferentiated pleomorphic sarcoma of the lateral femur	90	70*	3.6	1.23
4	SMARCA4 deficient undifferentiated uterine sarcoma (SDUS)	80	N.D.	39.4	56.3^{+}
9	Breast cancer of no special type (NST)	80	N.D.	3.2	1.65
13	Endometrioid carcinoma grade 2	80	N.D.	77.3	58.5^{+}
14	Moderately differentiated adenocarcinoma of the colon	50	2	44.7	20.34^{+}
16	Moderately differentiated adenocarcinoma of the colon	80	11	48.6	62.6^{+}
17	Seraseq TMB Mix Score 7	66	N.D.	7‡	N.D.
18	Seraseq TMB Mix Score 13	66	N.D.	13‡	N.D.
19	Seraseq gDNA HRD Low-Positive	66	58‡	N.D.	N.D.
20	Seraseq gDNA HRD Negative	66	33‡	N.D.	N.D.

Fig 1

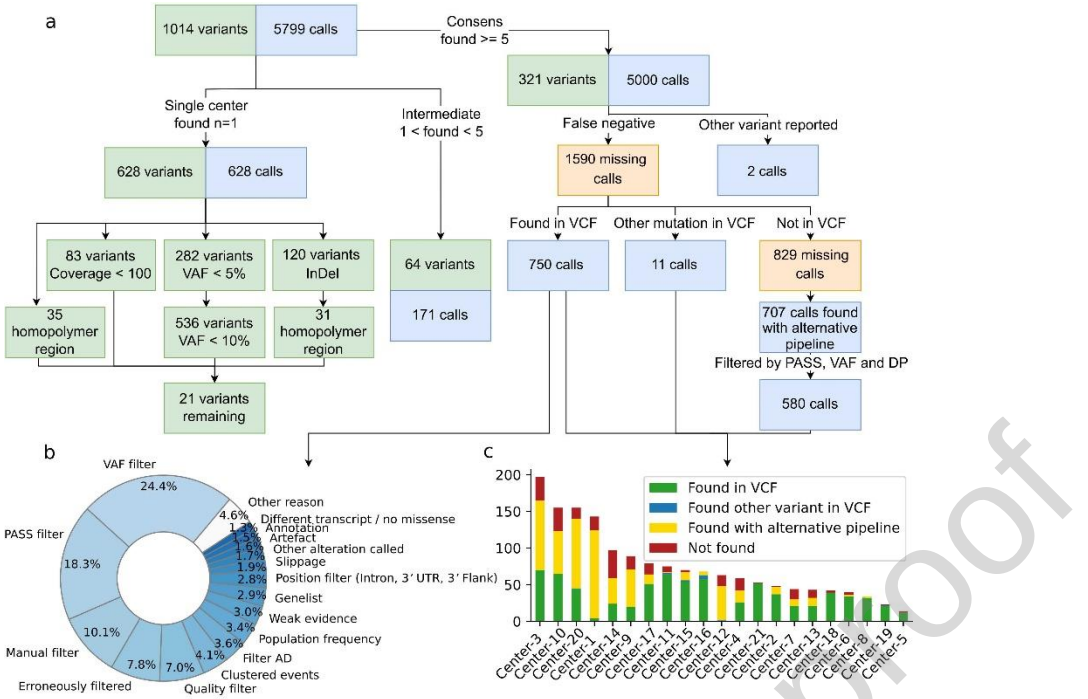


Fig 2

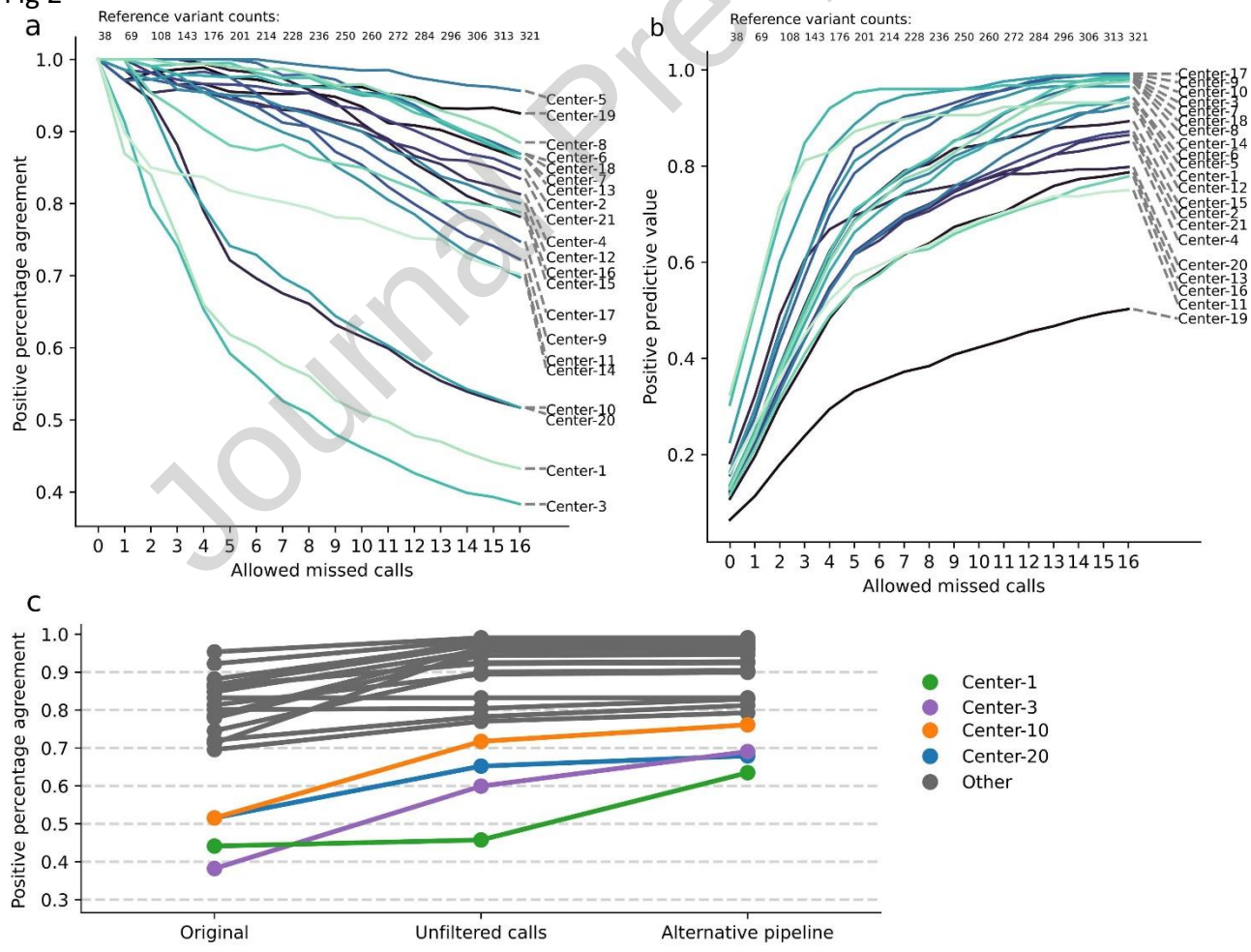


Fig 3

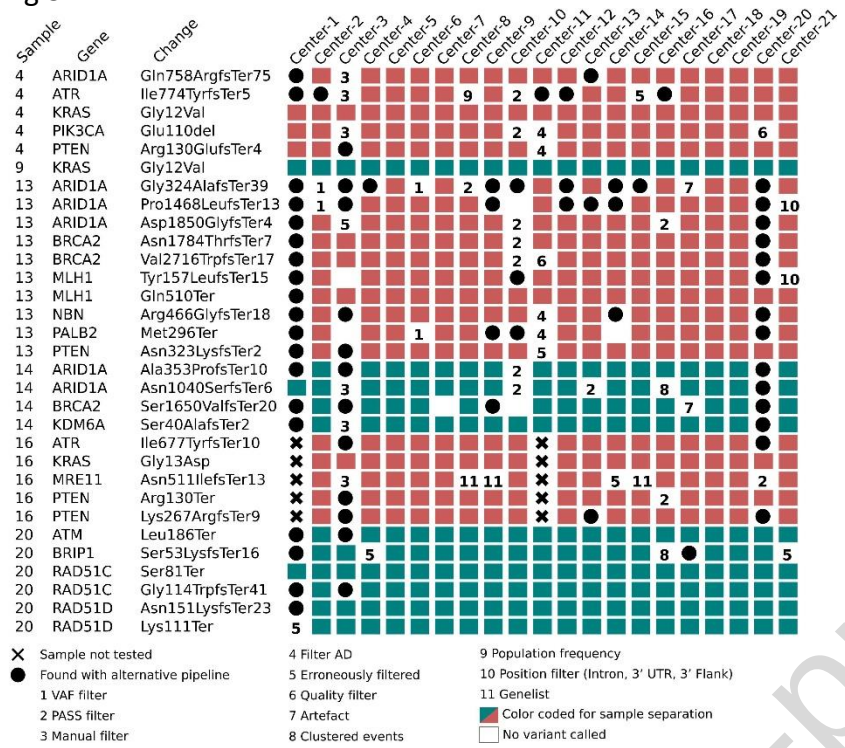


Fig 4

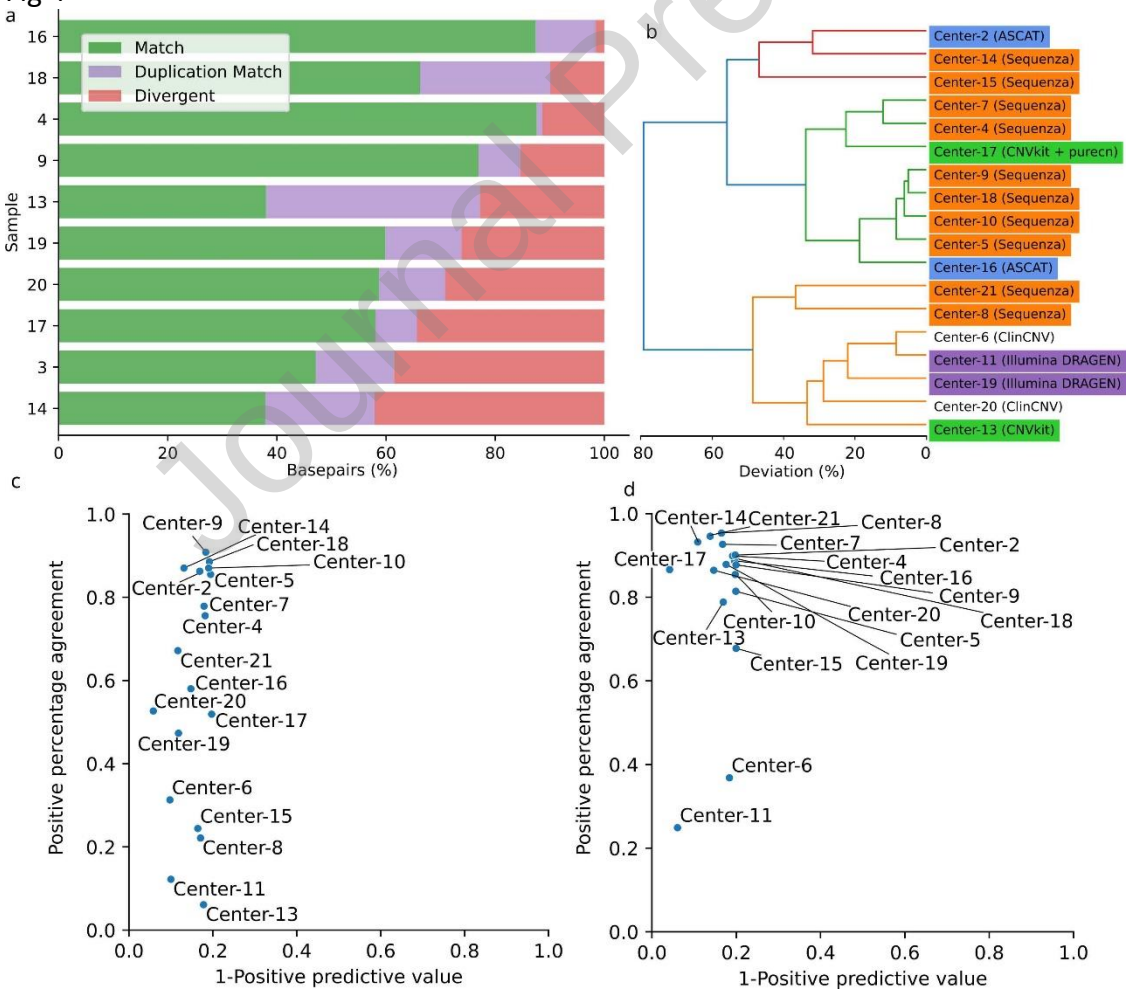
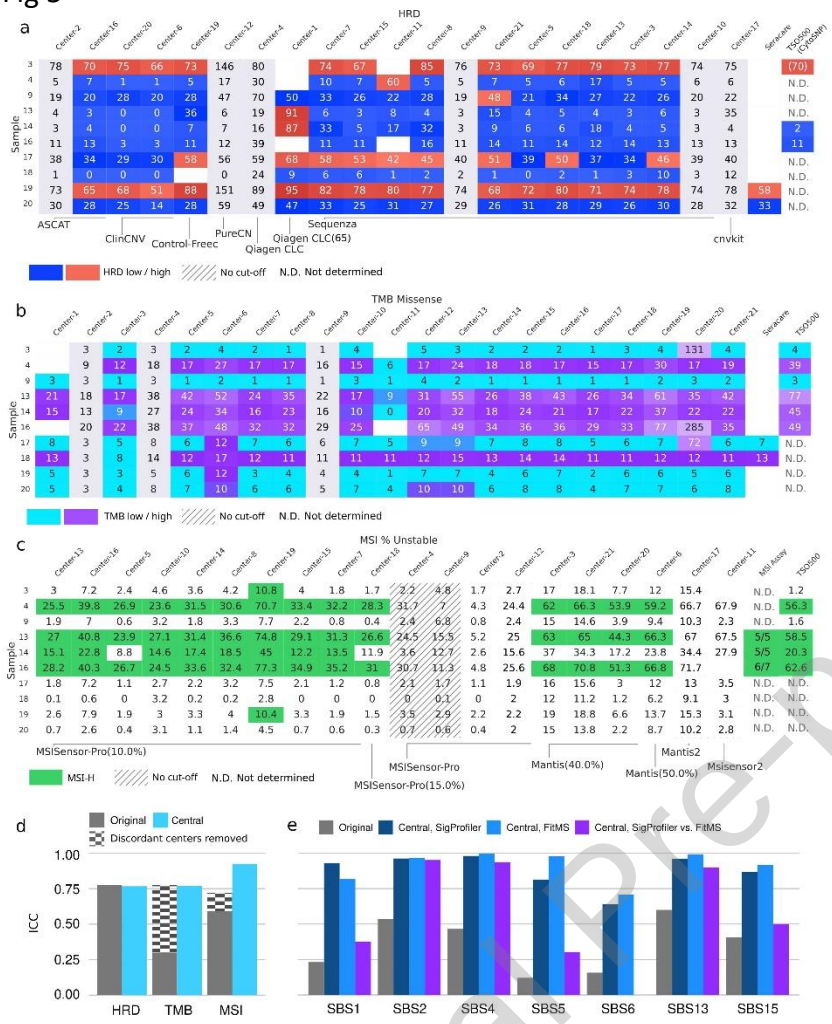


Fig 5



Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

MMT reports speaker and travel Expenses from Twist. JS reports speaker honoraria from DLS, Molecular Health, AstraZeneca and Biocartis, outside the submitted work. UL reports speaker fees from AstraZeneca, GSK, Novartis, Menarini, advisory board from AstraZeneca and Novartis. DH reports speaker honorary AstraZeneca, adboard BMS, WD speaker honoraries BMS & Novartis. SMB reports speaker honoraria, advisory board fees and research grants from AstraZeneca, Daiichi, Menarini, Novartis, Roche, BMS, Pfizer, Bayer, MSD, Merck, Amgen, Molecular Health, Targos, DLS, Janssen, GSK, QuIP, outside the submitted work. SL reports research grant from BMS, advisory board/speaker invitation from AstraZeneca, Eli Lilly, Roche and Takeda outside of this work. NTG reports research support from Janssen-Cilag and Advisory Boards from Janssen-Cilag, AstraZeneca, Daiichi-Sankyo and BMS outside the submitted work. WW reports research grants from Roche, MSD, BMS and AstraZeneca. Advisory board, lectures and speaker bureau fees from Roche, MSD, BMS, AstraZeneca, Pfizer, Merck, Lilly, Boehringer, Novartis, Takeda, Bayer, Janssen, Amgen, Astellas, Illumina, Eisai, Siemens, Agilent, ADC, GSK und Molecular Health. SO received reimbursement for travel expenses and payment for conference presentations from Illumina Inc. and Oxford Nanopore Technologies. CS reports research funding from BMS Stiftung Immunonkologie and institutional grants from Illumina outside the submitted work. CPS reports an investigator-initiated grant from Illumnia outside of the submitted work. PS reports grants from Inctye, BMS, Gilead, Falk, speakers bureau/advisory board from MSD, BMS, AstraZeneca, Incyte, Astellas, Janssen, Eisai, Amgen, Boehringer Ingelheim. DK reports personal fees for speaker honoraria from AstraZeneca, and Pfizer, personal fees for Advisory Board from Bristol-Myers Squibb, outside the submitted work. NP reports speaker fees from Novartis, Bayer, Roche, AstraZeneca, Illumina, BMS, MSD, PGDX/Labcorp, advisory board from Novartis, Lilly, Roche, Janssen, travel expenses from Novartis, AstraZeneca, Illumina, BMS, MSD, PGDX/Labcorp, Research grants from Illumina. JB reports grants from German Cancer Aid and consulting from MSD, outside the submitted work. AS reports participation in Advisory Board/Speaker's Bureau

Highlights

· WES analysis of 10 samples at 21 diagnostic centers resulting in 420 datasets

· Inter-center concordance of somatic and germline variant calling was high

· Most variant calling discordances were explainable by different variant filtering

· Copy number alteration calling was challenging and requires further standardization

· Complex biomarkers were mostly concordant, even using different bioinformatic tools