SUPPLEMENTARY METHODS

Extraction of nucleic acids

DNA and RNA from tumor specimens and DNA from matched blood samples were isolated using the AllPrep DNA/RNA/Protein Mini Kit or the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). For formalin-fixed and paraffin-embedded (FFPE) samples, the Generead DNA FFPE Kit or the AllPrep DNA/RNA FFPE Kit (Qiagen) were used. DNA from saliva samples was isolated using the prepIT-L2P kit (Genotek). Quality control and quantification were done using a Qubit Fluorometer (Life Technologies), a 4200 or 2200 TapeStation system (Agilent), and a 2100 Bioanalyzer system (Agilent).

Library preparation and target capture for whole-exome sequencing

For whole-exome sequencing (WES) library preparation, 1.5 µg genomic DNA were fragmented to 150-200 base pair (bp; paired-end) insert size with a Covaris S2 device, and 250 ng of Illumina adapter-containing libraries were hybridized with exome baits at 65°C for 16 hours. Exome capturing was performed using SureSelect Human All Exon in-solution capture reagents (Agilent). V4 with UTRs was used for the first 12 cases until May 2014 and V5 afterward. In case RNA was pooled in for sequencing, V5 without UTRs was used to reach a minimum average coverage of 80x for the tumor and 50x for the control. V5 with UTRs was used when DNA was sequenced alone.

Library preparation for whole-genome sequencing

Whole-genome sequencing (WGS) libraries were prepared using the TrueSeq Nano Library Preparation Kit (Illumina) following the manufacturer's instructions.

Library preparation for RNA sequencing

RNA sequencing (RNA-seq) libraries were prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina). A transition from the unstranded to the stranded protocol took place in February 2017. Briefly, mRNA was purified from 1 µg total RNA using oligo(dT) beads, poly(A)+ RNA was fragmented to 150 bp and converted into cDNA, and cDNA fragments were end-repaired, adenylated on the 3' end, adapter-ligated, and amplified with 12 cycles of PCR.

The final libraries were validated using a Qubit 2.0 Fluorometer (Life Technologies) and a Bioanalyzer 2100 system (Agilent).

Whole-exome, whole-genome, and RNA sequencing

Paired-end sequencing (2 x 151 bp) was performed with HiSeq X-Ten instruments (Illumina). Two lanes, each of tumor and control, were sequenced, yielding an average coverage of at least 70x for WGS cases. Paired-end sequencing (2 x 101 bp) was carried out with various HiSeq instruments (Illumina), in most cases pooling two patients' samples on one lane. Over time, we switched from HiSeq 2000 to HiSeq 2500 (mostly in rapid mode) and finally to HiSeq 4000 in April 2016. From January 2017, RNA was sequenced separately with dual indexing in pools of three samples per HiSeq 4000 lane or multiplexed over several lanes to prevent adapter hopping.

Alignment and mapping of whole-exome and whole-genome sequencing data

WES reads from HiSeq 2000 and HiSeq 2500 instruments were mapped to the 1000 Genomes Phase 2 assembly of the human reference genome (NCBI build 37.1, downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence) using a hardware-accelerated version of BWA (1) aln version 0.6.2 with default parameters and bwa sampe with maximum insert size set to 1,000 bp. BAM files were sorted with SAMtools (2) version 0.1.19, and PCR duplicates were marked with Picard tools (http://broadinstitute.github.io/picard/) version 1.90. WES reads from HiSeq 4000 instruments and WGS reads until August 2017 were aligned with BWA mem version 0.7.15 with -T 0 and all other parameters set to default. BAM files were sorted, and duplicates were marked with novosort version 1.03.07 using the DKFZ High-Throughput Sequencing Unit pipeline. Later samples were processed with the DKFZ One Touch Pipeline (OTP) (3) using BWA mem 0.7.10 and Sambamba (4) (http://lomereiter.github.io/sambamba/) version 0.4.6 and 0.5.9, respectively, for sorting and duplicate marking while merging the per-lane BAM files.

Variant calling

Single-nucleotide variants. For the detection of single-nucleotide variants (SNVs), we applied our in-house analysis pipeline based on SAMtools mpileup and bcftools version 0.1.19 with parameter adjustments to allow for

calling of somatic variants and heuristic filtering as previously described (5-7). Briefly, SAMtools mpileup is called on the tumor sample BAM file with parameters RE -q 20 -ug to consider only reads with a minimum mapping quality of 20 and bases with a minimum base quality of 13. The output is piped to bcftools view, which by using parameters -vcgN -p 2.0, reports all positions containing at least one high-quality non-reference base. From these initial putative SNV calls, we retain the ones with at least five variant reads and a variant allele frequency (VAF) of at least 5%. Any variant call supported by reads from only one strand is discarded if one of the Illumina-specific error profiles (8) occurs in a sequence context of ten bases around the putative SNV. For categorizing variants as germline or somatic, a pileup of the bases in the tumor-matched control sample is generated for each putative SNV position by SAMtools mpileup with parameters -Q 0 -q 1, considering uniquely mapping reads and refraining from a restriction on base quality. For high-confidence somatic SNVs, the coverage at the position in the control must be at least ten, and less than 1/30 of the control bases may support the variant observed in the tumor. Our confidence scoring scheme punishes overlap with genomic regions known to be prone to artifacts (9,10), which we annotate by several BED files downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTables). Starting with a value of 10, three points are subtracted for variants located in regions of low mappability (wgEncodeCrgMapabilityAlign100mer) or the hiSeqDepthTopPct1Pct track, Encode DAC Blacklisted Regions, or Duke Excluded Regions. A punishment of two points results from overlap with any two of the following features: short tandem repeats, simple repeats, low complexity, satellite repeats, or segmental duplications. Additionally, combinations of low VAF and a low number of variant-supporting reads are punished heuristically. If the confidence score falls below 1, it is set to 1. High-confidence SNVs are defined as having a score of at least 8 for WGS; the confidence cut-off for WES was 7 in the original version, where the presence of variant reads in the control led to a reduction of the score. At the time of switching to the OTP in November 2018, the cut-off was increased to 8. The pan-cancer workflow (11) for SNVs applied in OTP and for genomes requires a minimum mapping quality of 30, and calling of small insertions and deletions (indels) is deactivated. Here, all positions with at least three variant reads are called, but less than five variant reads lead to reducing the confidence score by three. To remove artifacts such as guanine oxidation, PCR and sequencing strand bias annotation was introduced. Due to the inherent sequencing strand bias in WES, only PCR bias leads to a reduced confidence score. In addition to regular confidence scoring, we introduced rescoring for removing artifacts prevalent in DNA from FFPE samples and in sequencing data from Illumina X-Ten systems. With standard confidence scoring, we identified recurrent high-confidence SNVs in *ZNF*

and *KRTAP* genes, *AHNAK2*, *FLG*, *MUC17*, and *MUC21*. Close investigation revealed that these SNVs frequently overlap with simple tandem repeats and have severe mapping quality and/or base quality biases in the PV4 field of the output of bcftools. We assume that DNA damage in FFPE samples and the very lenient mapping of BWA mem, especially for long reads, allows for mismapping of reads that would otherwise be clipped or not mapped. Consequently, we implemented additional penalties for simple tandem repeats and PV4 biases for WGS samples and finally also applied them routinely to WES samples since December 2017. In the rescoring, high-confidence SNVs may not overlap with an internal list of positions where SNVs are called with allele frequencies <10% in control samples (present in more than 18 of 624 WGS samples compiled from our International Cancer Genome Consortium [ICGC] and MASTER cohorts) or an entry in the pan-cancer bias-prone alleles blacklist (11). At the same time, the former "double punishment" of SNVs that have less than six variant reads and <10% VAF was relaxed to less than five reads and <5% VAF to allow improved calling in samples with low tumor cell content (TCC). In addition, somatic SNVs with a VAF less than five times the VAF in the control are set to lower confidence. Annotation was initially performed with RefSeq using the ANNOVAR (12) versions from September 2013 and August 2014, GENECODE version 19, and the ANNOVAR version from November 2014 in the pan-cancer OTP workflow. Annotations include information about nucleotide and amino acid changes and COSMIC (https://cancer.sanger.ac.uk/cosmic) and dbSNP (https://www.ncbi.nlm.nih.gov/snp) entries from the current version at that timepoint. For reporting, somatic, non-silent coding variants (i.e., non-synonymous, stop-gain, stop-loss, or splicing alterations within two bp of exon boundaries) of high confidence were selected.

Indels. Short indels were first identified using SAMtools mpileup and from February 2014 with Platypus (13) version 0.5.2; parameters genIndels=1, genSNPs=0, ploidy=2, nIndividuals=2 by providing the tumor and control BAM files. The OTP indel workflow uses Platypus version 0.8.1 with default parameters. To be of high confidence, somatic calls (control genotype 0/0) were required to either have the Platypus filter flag PASS or pass custom filters allowing for low VAF using a scoring scheme similar to that applied on SNVs. Candidates with the badReads flag, alleleBias or strandBias if the VAF was less than 10%, or more than two of the remaining flags were discarded. Additionally, combinations of Platypus non-PASS filter flags, bad quality values, low genotype quality, very low variant counts in the tumor, and presence of variant reads in the control were not tolerated. Indels were annotated with ANNOVAR, and somatic high-confidence indels falling into a coding sequence or splice site were extracted. For WES, we use a minimum confidence of 8. It is relaxed to 7 for WGS because the platypus filter flag "allele bias",

which means <10% VAF, often co-occurs with other punishments. This may lead to discarding potentially important indels with a low variant count that may result from the relatively lower coverage of WGS compared to WES, especially for samples with low tumor purity.

Germline SNVs and indels. From August 2015, we investigated germline SNVs and indels in a set of cancer-predisposing genes. Variants must have at least 30% VAF in the control sample, may not be annotated as COMMON in dbSNP, and must have a frequency <0.1 in the 1000 Genomes reference. SNVs with ≥3 homozygous or ≥40 heterozygous individuals in ExAC (14) version r0.3.nonTCGA.sites are filtered out. Additionally, we use a list of 280 local controls. If the same variant is present in more than 50 individuals, it is also discarded. For indels, the cut-off is 10 in the local controls and 50 in ExAC when there is a unique exact match; if more entries or a different indel are present at the position, the indel is only annotated and not removed. Recurrent artifacts in WGS, especially the *MSH2* splice site deletion (chr2:47641558 rs587779194 GT>G), which is frequently assigned the badReads filter flag by Platypus, but in other samples HapScore and/or QD instead, were manually removed until finally requiring the PASS flag, thus discarding confidence 7 for germline indels. Previous allogeneic transplantation in patients with solid tumors results in donor leucocytes in the tumor tissue, making the donor's polymorphisms appear as somatic SNVs. Furthermore, blood drawn after transplantation may contain mainly donor DNA so that the patient's polymorphisms are not detected, making them appear somatic. Thus, borrowing from the germline variant filtering strategy, a polymorphism filter is applied for patients that received an allogeneic bone marrow transplant.

Structural variants. We searched for structural variants (SVs) with CREST (15), which was replaced by our in-house tool SOPHIA (https://bitbucket.org/utoprak/sophia/src/master/) for WGS in July 2017 and, after intense evaluation, for WES in July 2018. While CREST uses soft-clipped reads assembled with CAP3 (16) and mapped to the reference genome with BLAT (17) for identifying SV breakpoints, the SOPHIA algorithm is based on the supplementary alignments from BWA mem and discordantly mapping mates. It uses a panel of 3,261 normal controls from ICGC and MASTER for filtering artifacts and common germline SVs. A single split read is sufficient to report a high-confidence SV if there are two discordant mates. The annotation of deletion or duplication does not consider coverage but only the direction of reads. A somaticity score is reported as the number of times the same breakpoint has been called in the background cohort of normals; thus, a high number means that the SV probably has been missed in the matching control sample. Germline events are only reported if they occur in less than 3% of the background cohort. A germline event with low read support may indicate that there is tumor DNA in the control.

For all events, overlapping and adjacent genes and the closest cancer genes are annotated with BEDtools. SOPHIA also reports potential RNA contamination.

Somatic copy number aberrations in WGS samples. For WGS, somatic copy number aberrations (sCNAs) were identified using our in-house pipeline ACEseq (18). To increase the detection of very small deletions, information on SVs was used as additional input. For simultaneous estimation of pairs of values for TCC and ploidy, the parameter space allowed for investigation contained the range 0.15-1.0 for TCC and the range 1.0-6.5 for ploidy. The ACEseq workflow includes the option of using a "fake control" for cases in which the matched control sample is too degraded to reliably call sCNAs. In that setting, while the BAF is calculated from the actual matched control, coverage bins are taken from a fixed normal sample of the same sex. For each possible combination of TCC and ploidy, absolute and allele-specific copy numbers and the decrease in heterozygosity (DH) were estimated segment-wise. Allele-specific copy numbers were calculated as total copy number (TCN) divided by two for balanced segments and as a function of coverage and B allele read counts in the case of imbalanced segments. The weighted mean distance of all segments to the next allowed integer copy number state was calculated for total and allele-specific copy numbers, where allowed means even TCN states for balanced segments and any integer copy number state for imbalanced segments and allele-specific copy numbers. TCC/ploidy combinations requiring negative copy number states or a DH >1 for any segment were excluded. Local minima in the weighted mean distance were considered possible TCC/ploidy solutions for the sample and were visually evaluated to select the optimal solution for each sample. Additionally, TCC was estimated from the VAF distribution of somatic SNVs, and sCNA- and SNV-based estimates were integrated into an informed decision for a TCC/ploidy pair. To detect sCNAs, segments with a TCN at least 0.7 above the selected ploidy were defined as gains, and segments with a TCN at least 0.3 below the selected ploidy were defined as losses. Segments with loss of heterozygosity (LOH), including copy number neutral ones, were also reported. Segments not fitting to an integer TCN were considered subclonal sCNAs. We call sCNAs shorter than three megabase pairs (Mbp) focal and term gains of a TCN over 2.5-fold the ploidy as amplifications. Since ACEseq, in contrast to CNVkit, does not assign genes to sCNAs, annotation was performed using the RefSeq gene from September 2013 with closestBed from the BEDtools package.

sCNAs in WES samples. For WES, sCNAs were initially analyzed by read depth plots and an in-house pipeline using the VarScan2 (19) copynumber and copyCaller modules. Regions were filtered for unmappable genomic stretches and merged by requiring at least 70 markers per called copy number event. Adjacent segments were merged

and filtered based on a log2 coverage ratio with a cut-off of 0.55 for gain and -0.55 for loss. Focal events within a smoothed segment were reported separately to identify amplifications and gains inside long deletions or vice versa. sCNAs were annotated with RefSeq genes using BEDTools. In December 2017, we switched to CNVkit (20). The VarScan2 pipeline often missed focal homozygous deletions of *CDKN2A* because the number of markers was below the cut-off for short regions, and thus such segments were filtered out as artifacts. Furthermore, in samples with low tumor purity where the fixed cut-offs are only reached for homozygous deletions and high-level amplifications, sCNAs with a lower log2 ratio could not be called. CNVkit version 0.8.6 was run with default parameter settings. Heterozygous single-nucleotide polymorphisms (SNPs) were determined as positions with an alternative allele fraction between 0.3 and 0.7 in the respective normal sample. Segments containing at least 20 heterozygous SNPs were further processed to infer sample ploidy and TCC along with allele-specific copy number estimates. Segments were classified as balanced or imbalanced according to the distribution of the frequencies of the alternative SNP alleles in the respective segments. If this distribution had a global maximum in the interval of 0.45-0.55, a segment was called balanced, and the remaining segments were further separated into two groups – ambiguous segments with one density peak outside the above interval and imbalanced segments with two peaks. Ambiguous segments were neglected in subsequent steps. For imbalanced segments, the mean BAF of all SNPs in the segment that were heterozygous in the germline was estimated using the allele with the higher read count as B allele. The mean read count of the B allele was calculated as product of the total coverage and the BAF of the respective segment. TCC and ploidy of a sample were estimated using a method adapted from ACEseq.

Quantification of genomic instability

A method to estimate homologous recombination deficiency, including large-scale state transition (LST) and telomeric allelic imbalance scores, was established in November 2017 based on the output of the CNVkit and ACEseq pipelines. Copy number profiles were smoothed to reduce oversegmentation caused by technical noise. All neighboring segments rounded to the same total and allele-specific copy number and did not deviate by more than 0.3 were merged. In addition to smoothing of similar segments, any segments <3 Mbp were merged with their more similar neighbor, as previously described (21). Any switch between copy number states of segments >10 Mbp that did not correspond to entire chromosome arms was counted as LST. Additionally, segments >15 Mbp that were less than a whole chromosome in length and corresponded to a LOH were counted for the LOH estimation.

Assessment of microsatellite instability

MSISensor (22) was applied to detect microsatellite instability (MSI). The list of homopolymers and microsatellites generated from the 1000 Genomes reference with the msisensor command "scan" comprises 33,386,244 loci. We run msisensor msi with a minimum required coverage of 15 reads for genomes and 30 for exomes in both tumor and control. MSISensor performs a standard $\chi 2$ test on the k-mer distribution of candidate microsatellite loci with sufficient coverage. If the k-mer distributions of tumor and control are significantly different, the microsatellite locus is classified as somatic MSI. The MSI score for a sample is the percentage of somatic instable microsatellites relative to the total number of microsatellites found in the control sample. According to published data (22), a score >3.5 implies MSI.

Supervised analysis of mutational signatures

We evaluated mutational signatures using the package YAPSA (Yet Another Package for Signature Analysis, http://bioconductor.org/packages/3.12/bioc/html/YAPSA.html). A linear combination decomposition of a sample's mutational catalog with predefined mutational signatures from the COSMIC database (http://cancer.sanger.ac.uk/cosmic/signatures) was computed by non-negative least squares (NNLS). When applied to WES cases, the mutational catalog was corrected for different occurrences of the triplet motifs between the whole genome and the target capture regions used for WES (function normalizeMotifs_otherRownames from YAPSA) before decomposition. To increase specificity, the NNLS algorithm was applied twice; after the first execution, only those signatures whose exposures, i.e., contributions in the linear combination, were higher than certain signature-specific cut-offs were kept, and the NNLS was rerun with the reduced set of signatures. As the detectability of the different signatures may vary, signature-specific cut-offs had been previously determined in a random operator characteristic analysis using publicly available data on mutational catalogs of 7,042 cancers (507 from WGS and 6,535 from WES) and mutational signatures from COSMIC. For WES cases, all somatic SNVs with confidence 7-10 were used.

Alignment of RNA sequencing data

RNA-seq reads were mapped with STAR (23) version 2.3.0e. For building the index, the 1000 Genomes reference sequence with GENECODE version 17 (https://www.gencodegenes.org/human/release_17.html) transcript annotations was used. For alignment, the following parameters were used: alignIntronMax 500000, alignMatesGapMax 500000, outSAMunmapped Within, outFilterMultimapNmax 1, outFilterMismatchNmax 3, outFilterMismatchNoverLmax 0.3, sjdbOverhang 50, chimSegmentMin 15, chimScoreMin 1, chimScoreJunctionNonGTAG 0, chimJunctionOverhangMin 15. The output was converted to sorted BAM files with SAMtools, and duplicates were marked with Picard tools version 1.90. From May 2016, we used STAR version 2.5 and for better chimeric junction calling adjusted the parameters with chimSegmentReadGapMax 3 --alignSJstitchMismatchNmax 5 -1 5 5. The OTP RNA workflow was used from July 2017 with a STAR index generated with the GENECODE version 19 gene model (https://www.gencodegenes.org/human/release_19.html) and duplicate marking with Sambamba. For alignment, the following parameters were used: --outFilterMultimapNmax 1 --outFilterMismatchNmax 5 --outFilterMismatchNoverLmax 0.3 --twopassMode Basic --twopass1readsN -1 --genomeLoad NoSharedMemory --chimSegmentMin 15 --chim ScoreMin 1 --chimScoreJunctionNonGTAG 0 --chimJunctionOverhangMin 15 --chimSegmentReadGapMax 3 --alignSJstitchMismatchNmax 5 -1 5 5 --alignIntronMax 1100000 --alignMatesGapMax 1100000 --alignSJDBoverhangMin 3 --alignIntronMin 20 --clip3pAdapterSeq TGGAATTCTCGGGTGCCAAGG.

Quantification of gene expression

Expression levels were determined per gene and sample as reads per kilobase (kb) of exon model per million mapped reads (RPKM), and RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) from September 12, 2013, was used as gene model. For each gene, overlapping annotated exons from all transcript variants in the genePred format were merged into non-redundant exon units with a custom Perl script. Nonduplicate reads with mapping quality above zero were counted for all exon units with coverageBed from the BEDtools package (24) version 2.16.2. To derive the RPKM value, read counts were summarized per gene and divided by the combined length of its exon units (in kb) and the total number of reads (in millions) as the sum of reads counted by coverageBed.

Identification of fusion genes from RNA sequencing data

Gene fusions were detected based on STAR chimeric reads, initially with a simple pipeline, from June 2016 using Arriba (https://github.com/suhrig/arriba), which removes recurrent alignment artifacts, transcript variants also observed in normal tissue, and reads with low sequence complexity as well as events with short anchors or breakpoints nearby or a low number of supporting reads relative to the overall number of predicted events in a gene.

Integration of information from the different layers of data

SNVs and indels were annotated with RNA information by generating a pileup of the DNA variant position in the RNA BAM file with SAMtools, parsed by a custom Perl script. SNVs with a very low VAF in the control might result from tumor DNA in the blood. In the case of leukemias, even a buccal swab or saliva sample may contain tumor cells. To "rescue" potential somatic SNVs mislabeled as germline by the above criteria, our in-house tool TiNDA (Tumor in Normal Detection Analysis) was used (25). Briefly, by detecting clusters of SNVs with low VAF in the control, but high VAF in the tumor, and annotation with the ExAC and GnomAD databases (https://gnomad.broadinstitute.org/), polymorphisms were discerned from rare germline variants and potential tumor-in-normal SNVs and indels. TiNDA rescue for SNVs was implemented for WGS samples in August 2017. From November 2015, somatic SNVs, indels, SVs, and sCNAs as well as germline SNVs and indels were integrated into one file per patient and augmented with additional annotations: dbNSFP version 2.9 (http://varianttools.sourceforge.net/Annotation/dbNSFP), from which scores of different functional impact prediction tools were used, ExAC (r0.3.nonTCGA.sites), 280 local controls, and several gene lists. The latter comprise the Cancer Gene Census (https://cancer.sanger.ac.uk/census) catalog as of December 2015 and the germline-predisposing gene list as well as manually curated and continuously updated lists of druggable lesions and other genes of interest (GOI), e.g., genes with resistance-causing mutations. Matching of the GOIs was done by gene ID. As part of an automated Clinical Bioinformatics Workflow, an Excel file is generated automatically for each patient, which includes tabs for the different variant types and a matrix that summarizes the alterations of the GOIs mentioned above. In case RNA information is present, RPKM and expression rank are annotated to the variants. Expression ranking is based on the first 148 RNA samples of the MASTER program. Genes from the Cancer Gene Census, the druggable lesions list, and the GOI list are reported as overexpressed if they have a fold change over two times the median of the comparison cohort, or, introduced as an additional criterion in October 2017 for genes that

are never two-fold upregulated, a z score above 2. Underexpression is reported for a fold change of 0.3 or less. For selected genes with alternative transcripts of particular interest, such as *CLDN18* (isoform 2 with exon 2 skipping) and *ALK* (alternative transcription initiation site in intron 19), per-exon RPKMs are reported in a separate file/tab.

Visualization of results for the molecular tumor board

For discussion in the molecular tumor board (MTB), variants of potential clinical relevance are presented in a digital slide show generated automatically from variant call files in VCF, BED(PE), and other tab-separated file formats. Variants are selected for presentation according to a predefined set of rules. Each rule specifies a gene and one or more conditions required for a gene to be selected. The list of genes was obtained from published datasets (26), the TARGET database (https://software.broadinstitute.org/cancer/cga/target), and the INFORM pilot study (27) and extended over time by additional candidate genes that have been shown to correlate with response to targeted therapies. Oncogenes are selected when affected by an SNV/indel, involved in a genomic rearrangement, strongly overexpressed, contained in a focal or high-level copy-number gain, or contained in a broad copy-number gain accompanied by moderate overexpression. Tumor suppressor genes are selected when affected by an SNV/indel, involved in a genomic rearrangement, extremely underexpressed, contained in a focal or homozygous deletion, or contained in a broad copy-number loss in conjunction with moderate underexpression. Genomic amplifications with an absolute copy number of more than 2.5x the base ploidy of a tumor are classified as high-level. Over-/underexpression is defined as an expression level in the top/bottom tenth percentile of the 148-sample reference subset of the MASTER cohort; extreme over-/underexpression is defined as an expression level in the top/bottom fifth percentile. In addition, a manually curated list of genes associated with germline variants predisposing to cancer is analyzed for germline SNVs and indels. Whenever necessary, variants for a recommendation in the MTB underwent a posteriori inspection with the Integrative Genomics Viewer (28).

Guidelines for clinical reassessment

Diagnoses were curated based on histopathologic reports and categorized using ICD-10-GM topography (version 2018) and ICD-O-3 morphology (third edition) codes. Basket assignments were made based on histopathologic reports and clinical considerations. Prior palliative systemic treatment was documented; neoadjuvant and adjuvant therapies were disregarded if the primary intention was curative and the patient had a progression-free interval >6

months. Clinical records were acquired before and three, six, 12, 18, and 24 months after the MTB. Outcome assessments were performed retrospectively by 10 specialists and clinical fellows in medical oncology or human genetics (D.Ha., C.E.H., C.H., P.H., A.J., S.K., A.M., L.M., L.R., and V.T.) using available clinical records. Documented implementation of recommended therapies led to further integrative evaluation of clinical outcomes. Treatment failure was assessed if a drug was administered for at least two cycles or over at least one month (for continuous treatments) and with at least 50% of the dose. Response was evaluated based on clinical and radiologic reports and categorized into five classes: complete response/no evidence of disease (CR), partial response (PR), mixed response (MR), stable disease for at least eight weeks (SD), and progressive disease (PD). MR, defined as concurrent response or stabilization of some and progression of other metastatic lesions accompanied by clinical benefit, was documented as an indication of biological drug efficacy and/or tumor heterogeneity. Response was not evaluated in cases with missing documentation and confounding therapeutic interventions, e.g., surgery or radiotherapy. Progression-free survival was determined for prior systemic therapies and therapies implemented according to MTB recommendations. Single drugs and individual components of combination therapies were assigned to one of seven molecular baskets based on the principal biomarker-drug relationship and/or the cellular pathways or processes involved. European Society for Medical Oncology Scale for Clinical Actionability of Molecular Targets tiers were added to each biomarker-drug combination, although the available clinical and preclinical evidence at the time of the MTB could only be approximated. Biomarkers were defined as any tumor- or patient-associated molecular alterations detected by WGS/WES and/or RNA-seq that informed clinical decision-making by the MTB. Biomarkers were evaluated in a binary system to allow clinical decisions based on a predictive biomarker's presence or absence.

REFERENCES

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **2009**;25(14):1754-60 doi 10.1093/bioinformatics/btp324.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009**;25(16):2078-9 doi 10.1093/bioinformatics/btp352.
3. Reisinger E, Genthner L, Kerssemakers J, Kensche P, Borufka S, Jugold A, *et al.* OTP: An automatized system for managing and processing NGS data. J Biotechnol **2017**;261:53-62 doi 10.1016/j.jbiotec.2017.08.006.
4. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics **2015**;31(12):2032-4 doi 10.1093/bioinformatics/btv098.
5. Jones DT, Hutter B, Jager N, Korshunov A, Kool M, Warnatz HJ, *et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. Nat Genet **2013**;45(8):927-32 doi 10.1038/ng.2682.
6. Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M, *et al.* Dissecting the genomic complexity underlying medulloblastoma. Nature **2012**;488(7409):100-5 doi 10.1038/nature11284.
7. Yaktapour N, Meiss F, Mastroianni J, Zenz T, Andrlova H, Mathew NR, *et al.* BRAF inhibitor-associated ERK activation drives development of chronic lymphocytic leukemia. J Clin Invest **2014**;124(11):5074-84 doi 10.1172/JCI76539.
8. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, *et al.* Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res **2011**;39(13):e90 doi 10.1093/nar/gkr344.
9. Pickrell JK, Gaffney DJ, Gilad Y, Pritchard JK. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. Bioinformatics **2011**;27(15):2144-6 doi 10.1093/bioinformatics/btr354.
10. Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, *et al.* Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat Biotechnol **2011**;30(1):61-8 doi 10.1038/nbt.2053.
11. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature **2020**;578(7793):82-93 doi 10.1038/s41586-020-1969-6.
12. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res **2010**;38(16):e164 doi 10.1093/nar/gkq603.
13. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Consortium WGS, *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet **2014**;46(8):912-8 doi 10.1038/ng.3036.
14. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res **2017**;45(D1):D840-D5 doi 10.1093/nar/gkw971.
15. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat Methods **2011**;8(8):652-4 doi 10.1038/nmeth.1628.

16. Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res **1999**;9(9):868-77 doi 10.1101/gr.9.9.868.
17. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res **2002**;12(4):656-64 doi 10.1101/gr.229202.
18. Kleinheinz K, Bludau I, Hübschmann D, Heinold M, Kensche P, Gu Z, *et al.* ACEseq – allele specific copy number estimation from whole genome sequencing. bioRxiv **2017**:210807 doi 10.1101/210807.
19. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res **2012**;22(3):568-76 doi 10.1101/gr.129684.111.
20. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol **2016**;12(4):e1004873 doi 10.1371/journal.pcbi.1004873.
21. Groschel S, Hubschmann D, Raimondi F, Horak P, Warsow G, Frohlich M, *et al.* Defective homologous recombination DNA repair as therapeutic target in advanced chordoma. Nat Commun **2019**;10(1):1635 doi 10.1038/s41467-019-09633-9.
22. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics **2014**;30(7):1015-6 doi 10.1093/bioinformatics/btt755.
23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics **2013**;29(1):15-21 doi 10.1093/bioinformatics/bts635.
24. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **2010**;26(6):841-2 doi 10.1093/bioinformatics/btq033.
25. Paramasivam N, Hubschmann D, Toprak UH, Ishaque N, Neidert M, Schrimpf D, *et al.* Mutational patterns and regulatory networks in epigenetic subgroups of meningioma. Acta Neuropathol **2019**;138(2):295-308 doi 10.1007/s00401-019-02008-w.
26. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet **2013**;45(10):1127-33 doi 10.1038/ng.2762.
27. Worst BC, van Tilburg CM, Balasubramanian GP, Fiesel P, Witt R, Freitag A, *et al.* Next-generation personalised medicine for high-risk paediatric cancer patients - The INFORM pilot study. Eur J Cancer **2016**;65:91-101 doi 10.1016/j.ejca.2016.06.009.
28. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, *et al.* Integrative genomics viewer. Nat Biotechnol **2011**;29(1):24-6 doi 10.1038/nbt.1754.